

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2003, 18:49:21 ; Search time 69.1132 Seconds
(without alignments)
4217.989 Million cell updates/sec

Title: US-09-965-553-1
Perfect score: 18
Sequence: 1 tggcgcgcgttgcacatg 18

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estln:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rnd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18	100.0	129	14	B0853899
2	18	100.0	166	17	AL771403
3	18	100.0	177	17	AL757122
4	18	100.0	178	17	AQ958189
5	18	100.0	199	17	BH243164
6	18	100.0	216	17	CNS00711

7	18	100.0	263	17	BH243803	BH243803	AUIRB177R
8	18	100.0	388	17	B62585	B62585	T22E21TF TA
9	18	100.0	401	17	BH244138	BH244138	AUIWA26TF
10	18	100.0	410	17	BH244101	BH244101	AUIWA38TF
11	18	100.0	415	17	BH243289	BH243289	AUIHF63TR
12	18	100.0	418	17	BH243156	BH243156	AUIHC93TR
13	18	100.0	419	17	BH252783	BH252783	SALK_0138
14	18	100.0	422	17	BH244105	BH244105	AUIWA21TR
15	18	100.0	424	17	BH244079	BH244079	AUIRB53TR
16	18	100.0	437	17	BH244067	BH244067	AUIRF34TR
17	18	100.0	439	17	BH244367	BH244367	AUIRF77TR
18	18	100.0	444	17	BH243576	BH243576	AUIRC53TR
19	18	100.0	446	17	BH235368	BH235368	AUIVA26TR
20	18	100.0	451	17	CNS00M07	CNS00M07	Arabidops
21	18	100.0	452	17	CNS00O1V	CNS00O1V	Arabidops
22	18	100.0	455	17	CNS00R39	CNS00R39	Arabidops
23	18	100.0	456	14	BQ989005	BQ989005	QGF16H19.
24	18	100.0	457	17	AQ958592	AQ958592	LERAV51TR
25	18	100.0	460	17	B95970	B95970	F21D2TFC IG
26	18	100.0	470	17	BH235524	BH235524	AUIVC40TF
27	18	100.0	471	17	CNS00MSM	CNS00MSM	Arabidops
28	18	100.0	478	17	BH243935	BH243935	AUIRA20TF
29	18	100.0	483	17	CNS00MGM	CNS00MGM	Arabidops
30	18	100.0	486	17	B23757	B23757	F16P22TR IG
31	18	100.0	493	17	BH243182	BH243182	AUIHD82TF
32	18	100.0	493	17	BH243582	BH243582	AUIRB65TF
33	18	100.0	493	17	BH243774	BH243774	AUIRB67TF
34	18	100.0	505	17	CNS00O0V	CNS00O0V	Arabidops
35	18	100.0	506	17	B27265	B27265	T3D23TF TAM
36	18	100.0	514	17	BH243704	BH243704	AUIRA27TR
37	18	100.0	520	17	CNS00R7E	CNS00R7E	Arabidops
38	18	100.0	521	17	CNS00N25	CNS00N25	Arabidops
39	18	100.0	522	17	BH244760	BH244760	AUIA68TR
40	18	100.0	524	17	BH243951	BH243951	AUIRD63TR
41	18	100.0	527	14	BQ851421	BQ851421	QGB15E20.
42	18	100.0	528	17	BH244809	BH244809	AUIVA24TF
43	18	100.0	537	17	BH235380	BH235380	AUIVA05TR
44	18	100.0	543	17	CNS00NS4	CNS00NS4	Arabidops
45	18	100.0	552	17	BH236544	BH236544	AUIFRA64TF

ALIGNMENTS

RESULT 1
B0853899/c
LOCUS
DEFINITION
OGB21M16.yg.abl OG_ABCDI lettuce salinas Lactuca sativa cDNA clone
ACCESSION
B0853899
VERSION
B0853899.1 GI:22239364
KEYWORDS
SOURCE
ORGANISM
Lactuca sativa.
Lactuca sativa.
Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids II; Asterales; Asteraceae; Lactuceae;
Lactuca.
1 (bases 1 to 129)
Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegler,J., Ellison
,P., Kolman,J.J., Slabaugh,M.S., Livingston,K., Zhou,Y., Lai,Z.,
Church,S., Jackson,L. and Bradford,K.
Lettuce and Sunflower ESTs from the Compositae Genome Project
http://compgenomics.ucdavis.edu/
Unpublished (2002)
Contact: Alexander Kozik [R.W.Michelmore]
Department of Vegetable Crops, R.W.Michelmore Lab
University of California at Davis (UCD)
Asmundson Hall, UCD, Davis, CA 95616, USA
Tel: 1-(530)-742-1742
Fax: 1-(530)-752-9659
Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]

vector pAC161. The lines contain one or more T-DNA insertions. The DNA fragment(s) resulting from the PCR were directly sequenced to determine the genomic sequence flanking the insertion. Sequences displaying significant similarity to the *A. thaliana* nuclear genome sequence were processed for submission. T-DNA derived sequences were removed.

BASE COUNT 54 a 29 c 42 g 52 t
ORIGIN

Query Match 100.0%; Score 18; DB 17; Length 177;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 TGGCGCCGTCGCAATTG 18
Db 34 TGGCGCCGTCGCAATTG 17

RESULT 4
LOCUS A0958189 178 bp DNA linear GSS 28-JAN-2000
DEFINITION LERAV83TF LERA Arabidopsis thaliana genomic clone LERAV83, DNA sequence.

ACCESSION A0958189
VERSION A0958189.1 GI:6785890
KEYWORDS GSS.

SOURCE thale cress.
ORGANISM Arabidopsis thaliana

REFERENCE 1 (bases 1 to 178)
AUTHORS Buell, C.R., Lin, X., Pal, G., Barnstead, M., Bowman, C., Uterbach, T., Feldblyum, T., Liang, F., Creasy, T., and Fraser, C.M.
TITLE Genomic survey sequencing of Landsberg erecta ecotype of Arabidopsis thaliana and identification of sequence-based polymorphisms

COMMENT Unpublished (2000)
CONTACT: Xiaoying Lin
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: atetigr.org
For additional information, see <http://www.tigr.org/cdb/at.at.html>
Seq primer: TF
Class: shotgun.

JOURNAL
COMMENT Location/Qualifiers

1..178
/organism="Arabidopsis thaliana"
/strain="Landsberg erecta"
/db_xref="taxon:3702"
/clone="LERAV83"
/clone.lib="LERA"
/note="Organ: leaf; Vector: PHOS1; Total genomic DNA was sheared to 0.9-1 kbp before ligation."
BASE COUNT 49 a 26 c 30 g 73 t
ORIGIN

FEATURES
source Location/Qualifiers

Query Match 100.0%; Score 18; DB 17; Length 178;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 TGGCGCCGTCGCAATTG 18
Db 77 TGGCGCCGTCGCAATTG 94

RESULT 5
LOCUS BH243164 199 bp DNA linear GSS 13-NOV-2001
DEFINITION AUHHC32TR AUH Arabidopsis thaliana genomic clone AUHHC32, DNA

sequence.

ACCESSION BH243164
VERSION BH243164.1 GI:16918951
KEYWORDS GSS.

SOURCE thale cress.
ORGANISM Arabidopsis thaliana

REFERENCE 1 (bases 1 to 199)
AUTHORS Town, C.D., Whitelaw, C.A., Pal, G., Van Aken, S.E., Uterbach, T.V., Feldblyum, T.V., and Fraser, C.M.
TITLE Survey sequencing of Arabidopsis thaliana BAC F1307

COMMENT Unpublished (2001)
CONTACT: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtownetigr.org
From Wash. U contlig 1003.
Seq primer: TR
Class: sheared ends.

FEATURES
source Location/Qualifiers

1..199
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="AUHHC32"
/clone.lib="AUH"
/note="Vector: PHOS2; Site 1: BstXI; 2-3 kb sheared BAC DNA inserted into PHOS2 using BstXI linkers"
BASE COUNT 55 a 42 c 34 g 68 t
ORIGIN

Query Match 100.0%; Score 18; DB 17; Length 199;
Best Local Similarity 100.0%; Pred. No. 93;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 TGGCGCCGTCGCAATTG 18
Db 154 TGGCGCCGTCGCAATTG 171

RESULT 6
LOCUS CNS00T11/c 216 bp DNA linear GSS 28-JUN-1999
DEFINITION Arabidopsis thaliana genome survey sequence T7 end of BAC T3023 of TAMU library from strain Columbia of Arabidopsis thaliana, genomic survey sequence.

ACCESSION AT089107
VERSION AT089107.1 GI:5290247
KEYWORDS GSS.

SOURCE Arabidopsis thaliana.

ORGANISM Arabidopsis thaliana
REFERENCE 1 (bases 1 to 216)
AUTHORS Salanoubat, M., Choise, N., Artiguenave, F., Brotlier, P., Wincker, P., Samsou, D., Saurin, W., Weissenbach, J. and Quetier, F.

JOURNAL Unpublished
COMMENT Direct Submission
TITLE Submitted (25-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
Web : www.genoscope.cns.fr

JOURNAL 2 (bases 1 to 216)
REFERENCE Genoscope.

FEATURES
source Location/Qualifiers

1..216
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"

```

/clone="T3D23"
/clone.lib="TAMU"
/note="end : 17"
BASE COUNT      82 a      40 c      40 g      54 t
ORIGIN
Query Match
Best Local Similarity 100.0%; Score 18; DB 17; Length 216;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0.
QY      1 TGGCGCCGTTGCCAATTG 18
      |||||||
Db      135 TGGCGCCGTTGCCAATTG 118

RESULT 7
BH243803
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 263)
Town,C.D., Whitelaw,C.A., Pal,G., Van Aken,S.E., Utterback,T.V.,
Feildbyum,T.V. and Fraser,C.M.,
Survey sequencing of Arabidopsis thaliana BAC F9022
Unpublished (2001)
Other-GSSs: AUIRB17TF
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
From Wash. U contig 1257.
Seq primer: 7R
Class: sheared ends.
FEATURES
Source
location/Qualifiers
1..263
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="AUIRB17"
/clone.lib="AUIR"
/note="Vector: pHOS2; Site 1: BstXI; 2-3 kb sheared BAC
DNA inserted into pHOS2 using BstXI linkers."
BASE COUNT      61 a      45 c      48 g      109 t
ORIGIN
Query Match
Best Local Similarity 100.0%; Score 18; DB 17; Length 263;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1 TGGCGCCGTTGCCAATTG 18
      |||||||
Db      136 TGGCGCCGTTGCCAATTG 153

RESULT 8
BH2585
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
Tnate cress.

```

```

ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi
1 (bases 1 to 388)
REFERENCE
Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Linher,K., Golden,K.
Berry,K., Granger,D., Suh,E., Wible,C., Adams,M.D. and Venter
,J.C.
A BAC End Sequence database for Identifying Minimal Overlaps in
Arabidopsis genomic Sequencing. Update 3
Unpublished (1997)
COMMENT
Contact: Steve Rounsley
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: rounsley@tigr.org
Seq primer: M13-21
Class: BAC ends
High quality sequence stop: 388.
FEATURES
source
1.388
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="T22F21"
/clone_1lb="TAMU"
/sex="hermaphrodite"
/notice="Vector: BelbacII; Site_1: HindIII; Site_2: HindIII
; Produced by Rod Wing"
BASE COUNT
143 a 71 c 77 g 97 t
ORIGIN
Query Match 100.0%; Score 18; DB 17; Length 388;
Best Local Similarity 100.0%; Pred. No. 1,3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TGGCGCCGTTGCCAATTG 18
|||||
DB 142 TGGCGCCGTTGCCAATTG 125
|||||
RESULT 9
BH244138 401 bp DNA GSS 13-NOV-2001
LOCUS
AUTMA26T6 AUTM Arabidopsis thaliana genomic clone AUTMA26, DNA
sequence.
ACCESSION
BH244138
VERSION
BH244138.1 GI:16920790
KEYWORDS
GSS.
SOURCE
thale cress.
ORGANISM
Arabidopsis thaliana.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi
1 (bases 1 to 401)
TOWN,C.D., Whitelaw,C.A., Pat,G., Van Aken,S.E., Uterback,T.V.,
Feldblum,T.V. and Fraser,C.M.
Survey sequencing of Arabidopsis thaliana BAC F13C23
Unpublished (2001)
Other GSSs: AUTMA26TR
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
From Wash. U contig 1440.
Seq primer: TP
Class: sheared ends.
FEATURES
source
1.401
Location/Qualifiers

```

```

/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="AU1WA26"
/clone_11b="AU1W"
/Note="Vector: PHOS2; Site_1: BstXI; 2-3 kb sheared BAC
DNA inserted into PHOS2 using BstXI linkers"
BASE COUNT      114 a      91 c      69 g      127 t
ORIGIN

Query Match      100.0%; Score 18; DB 17; Length 401;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 TGGCGCGGTGCCAATTG 18
Db      118 TGGCGCGGTGCCAATTG 135

RESULT 10
LOCUS   BH244101/c      410 bp      DNA      linear      GSS 13-NOV-2001
DEFINITION AU1WA38TR AU1W Arabidopsis thaliana genomic clone AU1WA38, DNA
ACCESSION BH244101
VERSION   BH244101.1 GI:16920714
KEYWORDS  GSS.
SOURCE    thale cress.
ORGANISM  Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 410)
Town,C.D., Whitelaw,C.A., Pai,G., Van Aken,S.E., Uterback,T.V.,
Feidlyum,T.V. and Fraser,C.M.
Survey sequencing of Arabidopsis thaliana BAC F13C23
Unpublished (2001)
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
From Wash. U contig 1440.
Seq primer: TR
Class: sheared ends.
FEATURES
    source
        1..410
            /organism="Arabidopsis thaliana"
            /strain="Columbia"
            /db_xref="taxon:3702"
            /clone="AU1WA38"
            /clone_11b="AU1W"
            /note="Vector: PHOS2; Site_1: BstXI; 2-3 kb sheared BAC
            DNA inserted into PHOS2 using BstXI linkers"
BASE COUNT      125 a      70 c      93 g      122 t
ORIGIN

Query Match      100.0%; Score 18; DB 17; Length 410;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 TGGCGCGGTGCCAATTG 18
Db      302 TGGCGCGGTGCCAATTG 285

RESULT 11
LOCUS   BH243289      415 bp      DNA      linear      GSS 13-NOV-2001
DEFINITION AU1HF63TR AU1H Arabidopsis thaliana genomic clone AU1HF63, DNA
sequence.
ACCESSION BH243289

```

```

VERSION BH243289.1 GI:16919186
KEYWORDS GSS.
SOURCE    thale cress.
ORGANISM  Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 415)
Town,C.D., Whitelaw,C.A., Pai,G., Van Aken,S.E., Uterback,T.V.,
Feidlyum,T.V. and Fraser,C.M.
Survey sequencing of Arabidopsis thaliana BAC F1307
Unpublished (2001)
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
From Wash. U contig 1003.
Seq primer: TR
Class: sheared ends.
FEATURES
    source
        1..415
            /organism="Arabidopsis thaliana"
            /strain="Columbia"
            /db_xref="taxon:3702"
            /clone="AU1HF63"
            /clone_11b="AU1H"
            /note="Vector: PHOS2; Site_1: BstXI; 2-3 kb sheared BAC
            DNA inserted into PHOS2 using BstXI linkers"
BASE COUNT      116 a      99 c      81 g      119 t
ORIGIN

Query Match      100.0%; Score 18; DB 17; Length 415;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 TGGCGCGGTGCCAATTG 18
Db      8 TGGCGCGGTGCCAATTG 25

RESULT 12
LOCUS   BH243156/c      418 bp      DNA      linear      GSS 13-NOV-2001
DEFINITION AU1HC93TR AU1H Arabidopsis thaliana genomic clone AU1HC93, DNA
sequence.
ACCESSION BH243156
VERSION   BH243156
KEYWORDS  GSS.
SOURCE    thale cress.
ORGANISM  Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 418)
Town,C.D., Whitelaw,C.A., Pai,G., Van Aken,S.E., Uterback,T.V.,
Feidlyum,T.V. and Fraser,C.M.
Survey sequencing of Arabidopsis thaliana BAC F1307
Unpublished (2001)
Other GSSs: AU1HC93TR
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
From Wash. U contig 1003.
Seq primer: TR
Class: sheared ends.
FEATURES
    source
        1..418
            /organism="Arabidopsis thaliana"

```

/strain="Columbia"
/db.xref="taxon:3702"
/clone="AUIHC93"
/clone_1lb="AUIH"
/note="Vector: PHOS2; Site_1: BstXI; 2-3 kb sheared BAC
DNA inserted into PHOS2 using BstXI linkers"

BASE COUNT

152 a 82 c 94 g 90 t

Query Match

Best Local Similarity 100.0%; Score 18; DB 17; Length 418;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TGGCGCCGTTGCCAATTG 18

Db 34 TGGCGCCGTTGCCAATTG 17

RESULT 13

BH252783

LOCUS BH252783 419 bp DNA linear GSS 28-NOV-2001

DEFINITION thaliana genomic clone SALK_013862, DNA sequence.

ACCESSION BH252783

VERSION BH252783.1 GI:17139761

KEYWORDS

SOURCE

thale cress.

ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

1 (bases 1 to 419)

Alonso,J.M., Llespe,T.J., Barajas,P., Chen,H., Cheuk,R., Gadinab,

C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P.

, Zimmermann,J. and Ecker,J.R.

A sequence-indexed library of insertion mutations in the

Arabidopsis genome

Unpublished (2001)

Contact: Joseph R. Ecker

Salk Institute Genomic Analysis Laboratory (SIGNAL)

The Salk Institute for Biological Studies

10010 N. Torrey Pines Road, La Jolla, CA 92037, USA

Tel: 858 453 4100 x1752

Fax: 858 558 6379

Email: jekers@salk.edu

This is single pass sequence recovered from the left border of

TDNA.

Class: TDNA tagged.

FEATURES

LOCATION/Qualifiers

1..419

/organism="Arabidopsis thaliana"

RESULT 14

BH244105/c

LOCUS BH244105 422 bp DNA linear GSS 13-NOV-2001

DEFINITION AUIWA21TR AUIW Arabidopsis thaliana genomic clone AUIWA21, DNA

sequence.

ACCESSION BH244105

VERSION BH244105.1 GI:16920722

KEYWORDS

SOURCE

thale cress.

ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

1 (bases 1 to 422)

Town,C.D., Whitelaw,C.A., Pal,G., Van Aken,S.E., Uterback,T.V.,

Feldblyum,T.V. and Fraser,C.M.

Survey sequencing of Arabidopsis thaliana BAC F13C23

Unpublished (2001)

Contact: Chris Town

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.

Tel: 301-838-3523

Fax: 301-838-0208

Email: cdtown@tigr.org

Class: sheared ends.

FEATURES

LOCATION/Qualifiers

1..422

/organism="Arabidopsis thaliana"

/db.xref="taxon:3702"

/clone_1lb="AUIWA21"

/note="Vector: PHOS2; Site_1: BstXI; 2-3 kb sheared BAC

DNA inserted into PHOS2 using BstXI linkers"

BASE COUNT

160 a 81 c 86 g 95 t

OY 1 TGGCGCCGTTGCCAATTG 18

Db 96 TGGCGCCGTTGCCAATTG 79

Query Match

Best Local Similarity 100.0%; Score 18; DB 17; Length 422;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 TGGCGCCGTTGCCAATTG 18

96 TGGCGCCGTTGCCAATTG 79

RESULT 15

BH244079 424 bp DNA linear GSS 13-NOV-2001

DEFINITION AUIR53TR AUIR Arabidopsis thaliana genomic clone AUIR53, DNA

sequence.

ACCESSION BH244079

VERSION BH244079.1 GI:16920670

KEYWORDS

SOURCE

thale cress.

ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

1 (bases 1 to 424)

Town,C.D., Whitelaw,C.A., Pal,G., Van Aken,S.E., Uterback,T.V.,

Feldblyum,T.V. and Fraser,C.M.

Survey sequencing of Arabidopsis thaliana BAC F9022

Unpublished (2001)

Contact: Chris Town

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.

Tel: 301-838-3523

Fax: 301-838-0208

Email: cdtown@tigr.org

From Wash. U contig 1257.
 Seq primer: TR
 Class: sheared ends.
 Location/Qualifiers

FEATURES
 source 1. 424
 /organism="Arabidopsis thaliana"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /clone="AU1RB53"
 /clone_lib="AU1R"
 /note="Vector: PHOS2; Site_1: BstXI; 2-3 kb sheared BAC
 DNA inserted into PHOS2 using BstXI linkers"
 BASE COUNT 94 a 83 c 83 g 164 t
 ORIGIN

Query Match 100.0%; Score 18; DB 17; Length 424;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 TGGGCGCGTGGCAATTG 18
 ||||||||||||||||
 Db 324 TGGGCGCGTGGCAATTG 341

Search completed: June 20, 2003, 22:39:12
 Job time : 71.1132 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2003, 19:31:16 ; Search time 75.6226 Seconds
(without alignments)
6927.165 Million cell updates/sec

Title: US-09-965-553-1

Perfect score: 18

Sequence: 1 tggcgccttgccatg 18

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 segs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*

- 1: gb_ba:*
- 2: gb_hg:*
- 3: gb_in:*
- 4: gb_ov:*
- 5: gb_om:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pl:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_sts:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vl:*
- 15: em_ba:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_mu:*
- 20: em_om:*
- 21: em_or:*
- 22: em_ov:*
- 23: em_pat:*
- 24: em_ph:*
- 25: em_pl:*
- 26: em_ro:*
- 27: em_sts:*
- 28: em_un:*
- 29: em_vl:*
- 30: em_hg_hum:*
- 31: em_hg_inv:*
- 32: em_hg_other:*
- 33: em_hg_mus:*
- 34: em_hg_pln:*
- 35: em_hg_rtd:*
- 36: em_hg_mam:*
- 37: em_hg_vrt:*
- 38: em_sy:*
- 39: em_hggo_hum:*
- 40: em_hggo_mus:*
- 41: em_hggo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	18	100.0	3539	8	AB073161	AB073161 Arabidops
2	18	100.0	18433	8	AC093090	AC093090 Arabidops
3	18	100.0	18813	8	T9E19	AF104920 Arabidops
4	18	100.0	32259	2	AC109921	AC109921 Arabidops
5	18	100.0	33329	8	AC007261	AC007261 Arabidops
6	18	100.0	35551	6	AX059461	AX059461 Sequence
7	18	100.0	36032	6	AX059500	AX059500 Sequence
8	18	100.0	38519	6	AX059548	AX059548 Sequence
9	18	100.0	39104	6	AX059479	AX059479 Sequence
10	18	100.0	40480	6	AX059456	AX059456 Sequence
11	18	100.0	42112	6	AX059497	AX059497 Sequence
12	18	100.0	42208	6	AX059477	AX059477 Sequence
13	18	100.0	47383	6	AX059459	AX059459 Sequence
14	18	100.0	47840	6	AX059452	AX059452 Sequence
15	18	100.0	48128	6	AX059469	AX059469 Sequence
16	18	100.0	48128	6	AX059474	AX059474 Sequence
17	18	100.0	48128	6	AX059513	AX059513 Sequence
18	18	100.0	48422	6	AX059509	AX059509 Sequence
19	18	100.0	50349	8	AC079028	AC079028 Arabidops
20	18	100.0	52616	2	AC090029	AC090029 Arabidops
21	18	100.0	54573	8	T6L9	AF147265 Arabidops
22	18	100.0	55345	8	AC007209	AC007209 Arabidops
23	18	100.0	57000	8	AB046428	AB046428 Arabidops
24	18	100.0	57000	8	AB046428	AB046428 Arabidops
25	18	100.0	58411	8	AB046430	AB046430 Arabidops
26	18	100.0	58411	8	AB046430	AB046430 Arabidops
27	18	100.0	60482	2	AC090030	AC090030 Arabidops
28	18	100.0	62916	8	AB046427	AB046427 Arabidops
29	18	100.0	64165	2	AC067965	AC067965 Arabidops
30	18	100.0	65316	8	F14C23	AF296828 Arabidops
31	18	100.0	66542	8	T10I18	AF262040 Arabidops
32	18	100.0	68352	8	T15F17	AF262042 Arabidops
33	18	100.0	69752	8	T5H22	AF096372 Arabidops
34	18	100.0	74265	8	AC068901	AC068901 Arabidops
35	18	100.0	76303	2	AC007183	AC007183 Arabidops
36	18	100.0	76911	2	AC016828	AC016828 Arabidops
37	18	100.0	77287	8	AB026642	AB026642 Arabidops
38	18	100.0	80279	2	AC011624	AC011624 Arabidops
39	18	100.0	81414	8	AB024037	AB024037 Arabidops
40	18	100.0	81806	8	AB062087	AB062087 Arabidops
41	18	100.0	81902	8	T17A2	AF160183 Arabidops
42	18	100.0	81902	8	T17A2	AF160183 Arabidops
43	18	100.0	83129	2	AC009529	AC009529 Arabidops
44	18	100.0	83530	8	AB062093	AB062093 Arabidops
45	18	100.0	84711	8	AP002067	AP002067 Arabidops

ALIGNMENTS

RESULT 1
AB073161/c 3539 bp DNA linear PLN 26-FEB-2002
LOCUS Arabidopsis thaliana DNA, chromosome 4 centromere region, BAC
DEFINITION clone:T5110.
ACCESSION AB073161
VERSION AB073161.1 GI:18149201
KEYWORDS SOURCE
ORGANISM Arabidopsis thaliana (cultivar:Columbia) DNA, clone:T5110.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1
Kumekawa,N., Hosouchi,T., Tsuruoka,H. and Kotani,H.

TITLE The size and sequence organization of the centromeric region of Arabidopsis thaliana chromosome 4

JOURNAL DNA Res 8 (6), 285-290 (2001)

REFERENCE 2 (bases 1 to 3539)

AUTHORS Kotani, H. and KumeKawa, N.

JOURNAL Submitted (16-OCT-2001) Hirokazu Kotani, Kazusa DNA Research Institute, Lab. Chromosome Research II, 1532-3 Yana, Kasarazu, Chiba 292-0812, Japan (E-mail:kotani@kazusa.or.jp, Tel:81-438-52-3920, Fax:81-438-52-3921)

FEATURES

Source

1. 3539
Location/Qualifiers
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="4"
/map="centromere"
/clone="T5110"
/note="this sequence was derived by PCR to cover the gap between clone T11G11 and clone F8H12"
1. 1000
/note="overlap to BAC T11G11, please refer for analysis and annotation"

misc_feature

1001..2539
/note="similarity to fragments of Athila retrotransposons"
2540..3539
/note="overlap to BAC F8H12, please refer for analysis and annotation"

BASE COUNT 1025 a 734 c 815 g 965 t

ORIGIN

Query Match 100.0%; Score 18; DB 8; Length 3539;
Best Local Similarity 100.0%; Pred. No. 62;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGCGCGGTGCCAATTG 18
1702 TGGCGCGGTGCCAATTG 1685

RESULT 2

AC093090 18433 bp DNA linear PLN 31-AUG-2001

LOCUS Arabidopsis thaliana chromosome 3 BAC F1C23 genomic sequence,

DEFINITION complete sequence.

AC093090

AC093090.1 GI:15145156

HTG.

SOURCE Arabidopsis thaliana.

ORGANISM Arabidopsis thaliana.

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis. 1 (bases 1 to 18433)

AUTHORS Town, C.D., Haas, B.J., Wu, D., Maitl, R., Hanick, L.T., Chan, A.P., Tallon, L.J., Rooney, T., Uterback, T.R., Vanaken, S.E., Feldblum, T.V., White, O., and Fraser, C.M.

JOURNAL Arabidopsis thaliana chromosome 3 BAC F1C23 genomic sequence unpublished

TITLE 2 (bases 1 to 18433)

AUTHORS Town, C.D., White, O., and Fraser, C.M.

JOURNAL Direct Submission

REFERENCE Submitted (09-AUG-2001) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA, cdtown@tigr.org

AUTHORS 3 (bases 1 to 18433)

JOURNAL Town, C.D. and Kaul, S.

TITLE Direct Submission

REFERENCE Submitted (31-AUG-2001) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA, cdtown@tigr.org

COMMENT Address all correspondence to: atetl@tigr.org

BAC clone F1C23 is from Arabidopsis thaliana chromosome 3

The orientation of the sequence is from SP6 to T7 end of the BAC clone.

Genes were identified by a combination of several methods: Gene prediction programs including Genscan+ (Chris Burge, <http://CCR-081.mt.edu/GENSCAN.html>), GenemarkHM (Mark Borodovsky, <http://genemark.biology.gatech.edu/Genemark/>), Glimmer4 (a variant of Glimmer3, see Minaela Perlea, <http://www.tigr.org/softlab/glimmer4.htm> and Geneslicer (Minaela Perlea and Steven Salzberg, <http://meritgate.tigr.org>), searches of the complete sequence against a peptide database and the plant EST database at TIGR (<http://www.tigr.org/tdb/tgi.shtml>). Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as unknown proteins. Genes without protein or EST similarity, that are predicted by more than two gene prediction programs over most of their length are annotated as hypothetical proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats are identified by repeatmasker (Arjan Smit, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>).

FEATURES

Source

1. 18433
Location/Qualifiers
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="3"
/map="7"
/clone="F1C23"
complement(365..405)
/rpt_family="AT-rich"
2852..5536
/gene="F1C23.1"
/note="pseudogene, putative Athila ORF1"
/pseudo
complement(16225..16754)
/gene="F1C23.2"
/note="similar to protein prospero GI:1346808 (Drosophila melanogaster)"
complement(join(16225..16384,16614..>16754))
/gene="F1C23.2"
complement(join(16226..16384,16614..16754))
/gene="F1C23.2"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAK96023.1"
/db_xref="GI:15383840"
/translation="MCMAYLVVFRYGCWSSNRGTEHEVRMPDASFSTLSOPWS PARGRWTIOSRHHYISLDKRTKPLKPSLINDIDHMKAFHSPLDANDTGRRLMT"

BASE COUNT 5393 a 3632 c 3666 g 5742 t

ORIGIN

Query Match 100.0%; Score 18; DB 8; Length 18433;
Best Local Similarity 100.0%; Pred. No. 74;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGCGCGGTGCCAATTG 18
2563 TGGCGCGGTGCCAATTG 2580

RESULT 3

T9E19/C 18813 bp DNA linear PLN 11-NOV-1998

LOCUS Arabidopsis thaliana BAC T9E19.

DEFINITION Arabidopsis thaliana BAC T9E19.

AC093090 AF104920

AC093090.1 GI:3859610

HTG.

SOURCE Arabidopsis thaliana.

ORGANISM Arabidopsis thaliana.

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

REFERENCE Rosidae: eurosids II; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 18813)
AUTHORS Washington University Genome Sequencing Center.
TITLE The A. thaliana Genome Sequencing Project.
JOURNAL Unpublished (1997)
REFERENCE 2 (bases 1 to 18813)
AUTHORS Stromwalt, C., Johnson, D. and Le, T.
TITLE The sequence of A. thaliana T9E19
JOURNAL Unpublished (1998)
REFERENCE 3 (bases 1 to 18813)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (08-NOV-1998) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT Submitted by: Genome Sequencing Center
Department of Genetics, Washington University,
St. Louis, MO 63108, USA
e-mail: twilson@watson.wustl.edu

MAPING: Clones were assigned to the YAC map by hybridization by M. Lodhi, Cold Spring Harbor Laboratories, and fingerprinted by M. Marra, WashU, to pick the best candidates for sequencing.

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone

NEIGHBORING COSMID INFORMATION:

The 5' clone is T2L5, 200 bp overlap; 3' clone is F7N22, 200 bp overlap. Actual start of this clone could not be found, the first known base of overlap is at 36622 of T2L5; actual end is at 76001 of F7N22.

NOTES:

Coding sequences below are predicted from computer analysis, using the program GeneFinder (P. Green and L. Hillier, ms in preparation).

FEATURES
SOURCE
Location/Qualifiers
1..18813
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="IV"
/map="5"
/clone="T9E19"
57..1256
/gene="T9E19.2"
join(57..893,969..1256)
/gene="T9E19.2"
/codon_start=1
/evidence-not_experimental
/protein_id="AAC72876.1"
/db_xref="GI:3859611"
/translation="MNAYSKPEYISDIENVLKGPMPOLLDSPFGELEKIPONKASE
NAKVLGICRQLVKKNEMIVFGGHPFRGLRRESITLGTGCKYKPKKVVNDVI
NVKPCSEVMNTLDERGCDVPTIADYSMTQEPESMGWKLALSLIIVGYVAA
HSNRPRTSKYVETKRLKFECKTPMGRTVSTRTGIRANFQVPAQQLIRQLVGS
YALGEPALQDLAFETIPSTAKLGPDDVLTNTPAFARSILRLAIPKIRLLECEA
ADEEVNVIKPADVNCPSLSMDDEVDPDVYDIALIDGQWODEDEWVGAYARP
KQVRPOLEETGKYKGCICQIFLVLHPETQLIATGILL"

gene
12757..15124
/gene="T9E19.1"
join(12757..13350,13635..13799,13898..14023,14096..14200,
14290..14391,14462..14521,14645..15124)
/gene="T9E19.1"
/codon_start=1
/evidence-not_experimental
/protein_id="AAC72877.1"
/db_xref="GI:3859612"

BASE COUNT 5547 a 3923 c 3965 g 5378 t
ORIGIN
Query Match 100.0%; Score 18; DB 8; Length 18813;
Best Local Similarity 100.0%; Pred. No. 74;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TGCCCGCGTTGCCAATTG 18
Db 11028 TGGCCCGCGTTGCCAATTG 11011

RESULT 4
AC109921 32259 bp DNA 1linear HTG 08-FEB-2002
LOCUS Arabidopsis thaliana chromosome unknown clone IGF-F9022 strain
DEFINITION Columbia, *** SEQUENCING IN PROGRESS ***, 7 unordered pieces.
ACCESSION AC109921 GI:18640653
VERSION AC109921
KEYWORDS HTG, HTGS_PHASE1.
SOURCE Arabidopsis thaliana.
ORGANISM Arabidopsis thaliana.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliopsida; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.

REFERENCE 1 (bases 1 to 32259)
AUTHORS Town, C.D., Haas, B., Tallon, L.J., Rooney, T., Ciecko, A.,
Uterback, T., Vanaken, S., Feldblum, T., White, O. and Fraser, C.M.
Arabidopsis thaliana 'IGF' BAC 'F9022' genomic sequence near marker
'?'

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 32259)
AUTHORS Town, C.D., White, O. and Fraser, C.M.
TITLE Direct Submission
JOURNAL Submitted (08-FEB-2002) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, cdtom@tigr.org

COMMENT
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1
* 3437: contig of 3437 bp in length
* 3438
* 3479: gap of unknown length
* 3480
* 5390: contig of 2511 bp in length
* 5391
* 6032: gap of unknown length
* 6033
* 9078: contig of 3046 bp in length
* 9079
* 9120: gap of unknown length
* 9121
* 11911: contig of 2791 bp in length
* 11912
* 11953: gap of unknown length
* 11954
* 15042: contig of 3089 bp in length
* 15043
* 15083: gap of unknown length

FEATURES	*	15084	24483: contig of 9400 bp in length
	*	24484	24524: gap of unknown length
	*	24525	32259: contig of 7735 bp in length.
FEATURES			Location/Qualifiers
source		1..32259	
		/organism="Arabidopsis thaliana"	
		/strain="Columbia"	
		/db_xref="taxon:3702"	
		/chromosome="unknown"	
		/clone="IGF-F9022"	
BASE COUNT		9587 a	6125 c 6878 g 9398 t 271 others
ORIGIN			
Query Match		100.0%;	Score 18; DB 2; Length 32259;
Best Local Similarity		100.0%;	Pred. No. 78;
Matches	18;	Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	1	TGGCGCCGTTGCCAATTG	18
DB	16753	TGGCGCCGTTGCCAATTG	16770
RESULT 5			
LOCUS	AC007261	33329 bp	DNA linear PLN 27-FEB-2002
DEFINITION	Arabidopsis thaliana chromosome 2 clone F16622 map g4532, complete sequence.		
ACCESSION	AC007261		
VERSION	AC007261.5	GI:20198225	
KEYWORDS	HTG.		
SOURCE	Arabidopsis thaliana.		
ORGANISM	Arabidopsis thaliana		
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.		
REFERENCE	1	(bases 1 to 33329)	
AUTHORS	Lin, X., Kaul, S., Shea, J.P., Fujii, C.Y., Shen, M., Vanaken, S.E., Bartsch, M.E., Mason, T.M., Bowman, C.L., Rongling, C.M., Beilto, M., Li, J., Carrera, A.J., Creasy, T.H., Buell, C.R., Town, C.D., Nierman, W.C., Fraser, C.M. and Venter, J.C.		
	Unpublished		
JOURNAL	2	(bases 1 to 33329)	
REFERENCE	Lin, X.		
AUTHORS	Direct Submission		
JOURNAL	Submitted (09-MAR-2000) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA		
REFERENCE	3	(bases 1 to 33329)	
AUTHORS	Town, C.D. and Kaul, S.		
JOURNAL	Direct Submission		
COMMENT	Submitted (27-FEB-2002) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA, cdtowne@igf.org		
FEATURES	On Apr 18, 2002 this sequence version replaced gi:6598734.		
source		Location/Qualifiers	
	1..33329		
	/organism="Arabidopsis thaliana"		
	/cultivar="Columbia"		
	/db_xref="taxon:3702"		
	/chromosome="2"		
	/map="g4532"		
	/clone="F16622"		
	/complement(11..482)		
	/note="overlap with BAC clone F15K19 (AC006429:1..482)."		
misc_feature	1		
	/note="2721 nt before this point were not included in the submitted sequence due to overlap with another BAC"		
repeat_region	complement(383..431)		
	/rpt_family="AT_rich"		
repeat_region	complement(2109..2135)		
	/rpt_family="AT_rich"		
repeat_region	complement(2211..2267)		
	/rpt_family="AT_rich"		
repeat_region	2482..2582		
	/rpt_family="AT_rich"		
repeat_region	2685..2724		
	/rpt_family="AT_rich"		
repeat_region	2951..2992		
	/rpt_family="AT_rich"		
repeat_region	complement(3184..3205)		
	/rpt_family="AT_rich"		
gene	5423..6109		
	/gene="At2g10930"		
	/note="F16622.1"		
mrna	join(5423..5657,5687..5861,5984..6109)		
	/gene="At2g10930"		
	join(5472..5657,5687..5861,5984..6072)		
	/gene="At2g10930"		
CDS	/codon_start=1		
	/product="unknown protein"		
	/protein_id="AAD28645.1"		
	/db_xref="GI:4733962"		
	/translation="MEVDSGSSGVYVNSKHYASLRADDEPETSLLSTLIDLVFRR LVLVLCYRVLMDOYQRTINLSQEVMEDEGRREKRPVPMKPIPVKRMKAPKRVDE DFKFLPQATYLLKQIQGGSGGFEFLRGVCTKDLSSNTFFYIF"		
	complement(6054..6076)		
	/rpt_family="AT_rich"		
	6215..7085		
misc_feature	/note="DNA sequence similar to Limpet transposon GB:U76697"		
	7256..7507		
gene	/gene="At2g10920"		
	/note="F16622.2"		
	7256..7507		
mrna	/gene="At2g10920"		
	7271..7442		
repeat_region	/rpt_family="(TGG)n"		
	7310..7507		
CDS	/gene="At2g10920"		
	/codon_start=1		
	/product="unknown protein"		
	/protein_id="AAD28646.1"		
	/db_xref="GI:4733963"		
	/translation="MNEEVYAEAEAVITRDLAEVVEKDYVAVREVAEEVTL EAFETPMGIVIKSLVLCGR"		
	complement(7921..7948)		
repeat_region	/rpt_family="AT_rich"		
gene	complement(8517..13115)		
	/gene="At2g10910"		
	/note="F16622.3; pseudogene, Tail-like non-LTR retroelement reverse transcriptase"		
	pseudo		
misc_feature	complement(11206..11773)		
	/gene="At2g10910"		
	/note="DNA sequence similar to non-LTR retrotransposon pseudogene for reverse transcriptase GB:LA7192"		
	pseudo		
repeat_region	complement(13353..13378)		
	/rpt_family="AT_rich"		
gene	complement(13763..16746)		
	/gene="At2g10900"		
	/note="F16622.4; similar to GB:AAA75253; some members of this protein family have a weak CCHC zinc fingers that is mostly from retroviral gag proteins (nucleocapsid)"		
mrna	complement(join<13763..14997,15212..15445,15881..16084,16536..216746)		
	/gene="At2g10900"		

Best Local Similarity 100.0%; Pred. No. 79;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGCGCCGTTGCCAATTG 18
DB 5453 TGGCGCCGTTGCCAATTG 5470

RESULT 8
LOCUS AX059548 38519 bp DNA linear PAT 17-JAN-2001
DEFINITION Sequence 281 from Patent WO0055325.
ACCESSION AX059548
VERSION AX059548.1 GI:12311655
KEYWORDS
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.

REFERENCE
1 (bases 1 to 38519)
AUTHORS Preuss, D., Copenhaver, G. and Keith, K.
TITLE Plant chromosome compositions and methods
JOURNAL Patent: WO 0055325-A 281 21-SEP-2000;
The University of Chicago (US)
FEATURES
source Location/Qualifiers
1..38519
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"

BASE COUNT 11817 a 7622 c 7434 g 11646 t
ORIGIN

Query Match 100.0%; Score 18; DB 6; Length 38519;
Best Local Similarity 100.0%; Pred. No. 80;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGCGCCGTTGCCAATTG 18
DB 238 TGGCGCCGTTGCCAATTG 255

RESULT 9
LOCUS AX059479 39104 bp DNA linear PAT 17-JAN-2001
DEFINITION Sequence 212 from Patent WO0055325.
ACCESSION AX059479
VERSION AX059479.1 GI:12311584
KEYWORDS
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.

REFERENCE
1 (bases 1 to 39104)
AUTHORS Preuss, D., Copenhaver, G. and Keith, K.
TITLE Plant chromosome compositions and methods
JOURNAL Patent: WO 0055325-A 212 21-SEP-2000;
The University of Chicago (US)
FEATURES
source Location/Qualifiers
1..39104
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"

BASE COUNT 11558 a 7680 c 7833 g 12033 t
ORIGIN

Query Match 100.0%; Score 18; DB 6; Length 39104;
Best Local Similarity 100.0%; Pred. No. 80;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGCGCCGTTGCCAATTG 18
DB 7762 TGGCGCCGTTGCCAATTG 7779

RESULT 10
LOCUS AX059456/c 40480 bp DNA linear PAT 17-JAN-2001
DEFINITION Sequence 189 from Patent WO0055325.
ACCESSION AX059456
VERSION AX059456.1 GI:12311561
KEYWORDS
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.

REFERENCE
1 (bases 1 to 40480)
AUTHORS Preuss, D., Copenhaver, G. and Keith, K.
TITLE Plant chromosome compositions and methods
JOURNAL Patent: WO 0055325-A 189 21-SEP-2000;
The University of Chicago (US)
FEATURES
source Location/Qualifiers
1..40480
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"

BASE COUNT 12306 a 7912 c 8765 g 11297 t 200 others
ORIGIN

Query Match 100.0%; Score 18; DB 6; Length 40480;
Best Local Similarity 100.0%; Pred. No. 80;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGCGCCGTTGCCAATTG 18
DB 35456 TGGCGCCGTTGCCAATTG 35439

RESULT 11
LOCUS AX059497 42112 bp DNA linear PAT 17-JAN-2001
DEFINITION Sequence 230 from Patent WO0055325.
ACCESSION AX059497
VERSION AX059497.1 GI:12311604
KEYWORDS
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.

REFERENCE
1 (bases 1 to 42112)
AUTHORS Preuss, D., Copenhaver, G. and Keith, K.
TITLE Plant chromosome compositions and methods
JOURNAL Patent: WO 0055325-A 230 21-SEP-2000;
The University of Chicago (US)
FEATURES
source Location/Qualifiers
1..42112
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"

BASE COUNT 12877 a 8525 c 8423 g 12287 t
ORIGIN

Query Match 100.0%; Score 18; DB 6; Length 42112;
Best Local Similarity 100.0%; Pred. No. 80;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGCGCCGTTGCCAATTG 18
DB 1458 TGGCGCCGTTGCCAATTG 1441

RESULT 12
LOCUS AX059477/c 42208 bp DNA linear PAT 17-JAN-2001
DEFINITION Sequence 210 from Patent WO0055325.
ACCESSION AX059477
VERSION AX059477.1 GI:12311582

KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
BASE COUNT
ORIGIN
Query Match
Best Local Similarity
Matches
OY
DB
RESULT 13
AX059459/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
BASE COUNT
ORIGIN
Query Match
Best Local Similarity
Matches
OY
DB
RESULT 14
AX059452/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 47840)

thale cress.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 42208)
Preuss,D., Copenhaver,G. and Keith,K.
Plant chromosome compositions and methods
Patent: WO 005325-A 210 21-SEP-2000;
The University of Chicago (US)
Location/Qualifiers
1. 42208
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
12522 a 8859 c 886 g 11961 t
100.0%; Score 18; DB 6; Length 42208;
100.0%; Pred. No. 80;
0; Mismatches 0; Indels 0; Gaps 0;
1 TGGCGCCGTTGCCAATTG.18
35889 TGGCGCCGTTGCCAATTG 35872
47383 bp DNA linear PAT 17-JAN-2001
Sequence 192 from Patent WO0055325.
AX059459
AX059459.1 GI:12311564
thale cress.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 47383)
Preuss,D., Copenhaver,G. and Keith,K.
Plant chromosome compositions and methods
Patent: WO 005325-A 192 21-SEP-2000;
The University of Chicago (US)
Location/Qualifiers
1. 47383
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
14209 a 9919 c 9688 g 13496 t 71 others
100.0%; Score 18; DB 6; Length 47383;
100.0%; Pred. No. 81;
0; Mismatches 0; Indels 0; Gaps 0;
1 TGGCGCCGTTGCCAATTG 18
7996 TGGCGCCGTTGCCAATTG 7979
47840 bp DNA linear PAT 17-JAN-2001
Sequence 185 from Patent WO0055325.
AX059452
AX059452.1 GI:12311557
thale cress.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 47840)

AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
BASE COUNT
ORIGIN
Query Match
Best Local Similarity
Matches
OY
DB
RESULT 15
AX059469
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
BASE COUNT
ORIGIN
Query Match
Best Local Similarity
Matches
OY
DB
Search completed: June 20, 2003, 23:23:56
Job time : 77.956 secs

Preuss,D., Copenhaver,G. and Keith,K.
Plant chromosome compositions and methods
Patent: WO 005325-A 185 21-SEP-2000;
The University of Chicago (US)
Location/Qualifiers
1. 47840
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
14095 a 10020 c 9567 g 14078 t 80 others
100.0%; Score 18; DB 6; Length 47840;
100.0%; Pred. No. 81;
0; Mismatches 0; Indels 0; Gaps 0;
1 TGGCGCCGTTGCCAATTG 18
35451 TGGCGCCGTTGCCAATTG 35434
48128 bp DNA linear PAT 17-JAN-2001
Sequence 202 from Patent WO0055325.
AX059469
AX059469.1 GI:12311574
thale cress.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 48128)
Preuss,D., Copenhaver,G. and Keith,K.
Plant chromosome compositions and methods
Patent: WO 005325-A 202 21-SEP-2000;
The University of Chicago (US)
Location/Qualifiers
1. 48128
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
15779 a 7963 c 8596 g 15790 t
100.0%; Score 18; DB 6; Length 48128;
100.0%; Pred. No. 81;
0; Mismatches 0; Indels 0; Gaps 0;
1 TGGCGCCGTTGCCAATTG 18
41528 TGGCGCCGTTGCCAATTG 41545

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2003, 18:46:16 ; Search time 10.3019 seconds
(without alignments)
3934.810 Million cell updates/sec

Title: US-09-965-553-1

Perfect score: 18

Sequence: 1 ttgcccgcgttgcacattg 18

Scoring table:

IDENTITY-NUC
Gapop 10.0' , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries.

Database :

N.Geneseq_101002:*
1: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	18	100.0	18	21	AAZ35269
2	18	100.0	149	21	AAZ35270
3	18	100.0	59590	21	AAZ35271
4	18	100.0	64415	21	AAZ35272
5	18	100.0	83390	21	AAZ35273
6	18	100.0	90336	21	AAZ35274
7	18	100.0	92584	21	AAZ35275
8	18	100.0	94895	21	AAZ35276
9	18	100.0	95223	21	AAZ35277

10	18	100.0	96583	21	AAZ35278	BAC containing rep
11	18	100.0	96988	21	AAZ35279	BAC containing rep
12	18	100.0	109973	21	AAZ35280	BAC containing rep
13	18	100.0	134499	21	AAZ35281	BAC containing rep
14	18	100.0	134499	21	AAZ35282	BAC containing rep
15	18	100.0	611590	21	AAZ35283	BAC containing rep
16	18	100.0	1082138	21	AAZ35284	BAC containing rep
17	17	94.4	40349	21	AAZ35285	BAC containing rep
18	17	94.4	40349	21	AAZ35286	BAC containing rep
19	17	94.4	151826	21	AAZ35287	BAC containing rep
20	16.4	91.1	50959	21	AAZ35288	BAC containing rep
21	16.4	91.1	64415	21	AAZ35289	BAC containing rep
22	16.4	91.1	72592	21	AAZ35290	BAC containing rep
23	16.4	91.1	79122	21	AAZ35291	BAC containing rep
24	16.4	91.1	83390	21	AAZ35292	BAC containing rep
25	16.4	91.1	86584	21	AAZ35293	BAC containing rep
26	16.4	91.1	90336	21	AAZ35294	BAC containing rep
27	16.4	91.1	103929	21	AAZ35295	BAC containing rep
28	15.4	85.6	548	21	AAZ35296	BAC containing rep
29	15.4	85.6	1713	21	AAZ35297	BAC containing rep
30	15.4	85.6	1713	21	AAZ35298	BAC containing rep
31	15.4	85.6	1713	21	AAZ35299	BAC containing rep
32	15.4	85.6	80450	21	AAZ35300	BAC containing rep
33	15.4	85.6	82588	21	AAZ35301	BAC containing rep
34	15.4	85.6	94895	21	AAZ35302	BAC containing rep
35	15.4	85.6	163319	21	AAZ35303	BAC containing rep
36	15.4	85.6	1082138	21	AAZ35304	BAC containing rep
37	15	83.3	412	21	AAZ35305	BAC containing rep
38	15	83.3	163319	21	AAZ35306	BAC containing rep
39	14.8	82.2	345	24	AAZ35307	BAC containing rep
40	14.8	82.2	417	21	AAZ35308	BAC containing rep
41	14.8	82.2	417	21	AAZ35309	BAC containing rep
42	14.8	82.2	573	24	AAZ35310	BAC containing rep
43	14.8	82.2	849	21	AAZ35311	BAC containing rep
44	14.8	82.2	986	21	AAZ35312	BAC containing rep
45	14.8	82.2	1089	24	AAZ35313	BAC containing rep

ALIGNMENTS

RESULT 1	AAZ35269	AAZ35269 standard; DNA; 18 BP.
AC	AAZ35269;	
DT	27-MAR-2000	(first entry)
DE	Plant retroelement primer binding site version 1.	
KW	Retroelement; retrovirus; transgenic plant; gene transfer; primer;	
KW	soybean; ss.	
OS	Glycine max.	
PN	W09960842-A2.	
PD	02-DEC-1999.	
PF	28-MAY-1999;	99WO-US11858.
PR	29-MAY-1998;	98US-0087125.
PR	28-MAY-1999;	99US-0087125.
PA	(WRIG/) WRIGHT D A.	
PA	(VOYT/) VOYTAS D F.	
PI	Wright DA, Voytas DF;	
DR	WPI; 2000-105586/09.	
XX	New nucleic acid molecules for imparting agronomically significant	

PT characters to plants, especially soybean
 XX
 PS Disclosure; Page 72; 118pp; English.
 CC This oligonucleotide represents a soybean retroelement primer
 CC binding site (version 1). The invention provides molecular tools
 CC in the form of retroelements and retroelement-containing vectors,
 CC cells and plants. Methods are provided for introducing the
 CC retroelements into cells, especially when the retroelement carries
 CC at least 1 agronomically-significant characteristic. In a
 CC preferred method, a helper cell line which expresses gag, pol and
 CC env sequences is used to enable transfer of a secondary construct
 CC which carries an agronomically-significant characteristic and has
 CC retroelement sequences that allow for replication and integration.
 CC Claimed isolated nucleic acid molecules comprise a nucleic acid
 CC sequence selected from a retroelement primer binding site,
 CC envelope, gag, integrase, reverse transcriptase, protease or
 CC RNase-H sequence (see AA35270-61). Also provided are plant
 CC retroviral particles that are used to transfer the nucleic acids
 CC into plant cells.
 CC
 SQ Sequence 18 BP; 2 A; 5 C; 6 G; 5 T; 0 other;
 Query Match 100.0%; Score 18; DB 21; Length 18;
 Best Local Similarity 100.0%; Pred. No. 8.3;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 TGGCGCCGTCGCAATTG 18
 Db 1 TGGCGCCGTCGCAATTG 18
 RESULT 2
 AA35270
 ID AA35270 standard; DNA; 149 BP.
 XX
 AC AA35270;
 XX
 DT 27-MAR-2000 (first entry)
 XX
 DE Plant retroelement primer binding site A.
 XX
 KM Retroelement: retrovirus; transgenic plant; gene transfer;
 XX primer binding site; ss.
 OS Unidentified.
 XX
 PN WO9960842-A2.
 XX
 PD 02-DEC-1999.
 XX
 PE 28-MAY-1999; 99WO-US11858.
 XX
 PR 29-MAY-1998; 98US-0087125.
 XX
 PR 28-MAY-1999; 99US-0087125.
 XX
 PA (WRIGHT) WRIGHT D A.
 XX (VOYT/) VOYTAS D F.
 XX
 PI Wright DA, Voytas DF;
 XX
 DR WPI; 2000-105586/09.
 XX
 PT New nucleic acid molecules for imparting agronomically significant
 PT characters to plants, especially soybean
 XX
 PS Disclosure; Page 115; 118pp; English.
 CC This oligonucleotide represents plant retroelement primer binding
 CC site A of the invention. The invention provides molecular tools
 CC in the form of retroelements and retroelement-containing vectors,
 CC cells and plants. Methods are provided for introducing the
 CC retroelements into cells, especially when the retroelement carries

CC at least 1 agronomically-significant characteristic. In a
 CC preferred method, a helper cell line which expresses gag, pol and
 CC env sequences is used to enable transfer of a secondary construct
 CC which carries an agronomically-significant characteristic and has
 CC retroelement sequences that allow for replication and integration.
 CC Claimed isolated nucleic acid molecules comprise a nucleic acid
 CC sequence selected from a retroelement primer binding site,
 CC envelope, gag, integrase, reverse transcriptase, protease or
 CC RNase-H sequence (see AA35270-61). Also provided are plant
 CC retroviral particles that are used to transfer the nucleic acids
 CC into plant cells.
 CC
 SQ Sequence 149 BP; 39 A; 23 C; 26 G; 61 T; 0 other;
 Query Match 100.0%; Score 18; DB 21; Length 149;
 Best Local Similarity 100.0%; Pred. No. 9.1;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 TGGCGCCGTCGCAATTG 18
 Db 66 TGGCGCCGTCGCAATTG 83
 RESULT 3
 AAF22281/C
 ID AAF22281 standard; DNA; 59590 BP.
 XX
 AC AAF22281;
 XX
 DT 20-MAR-2001 (first entry)
 XX
 DE BAC containing repeats from centromeres 1-4 #4.
 XX
 KM Centromere; microsome; vector; ds.
 XX
 OS Arabidopsis thaliana.
 XX
 PN WO200055325-A2.
 XX
 PD 21-SEP-2000.
 XX
 PE 17-MAR-2000; 2000WO-US07392.
 XX
 PR 18-MAR-1999; 99US-0125219.
 XX
 PR 01-APR-1999; 99US-0127409.
 XX
 PR 18-MAY-1999; 99US-0134770.
 XX
 PR 13-SEP-1999; 99US-0153584.
 XX
 PR 17-SEP-1999; 99US-0154603.
 XX
 PA (UYCH-) UNIV CHICAGO.
 XX
 PI Preuss D, Copenhaver G, Keith K;
 XX
 DR WPI; 2000-587529/55.
 XX
 PT Recombinant DNA construct comprising a plant centromere, useful for
 PT producing stably inherited microsomes which can serve as vectors for
 PT the construction of transgenic plant and animal cells
 XX
 PS Claim 102; Page 351-364; 1449pp; English.
 CC The present invention relates to a recombinant DNA construct of a plant
 CC (Arabidopsis thaliana) centromere. The constructs are useful for
 CC producing stably inherited microsomes which can serve as vectors for
 CC the construction of transgenic plant and animal cells expressing
 CC selected proteins such as hormones, enzymes, interleukins, clotting
 CC factors, cytokines, antibodies, and growth factors.
 CC
 SQ Sequence 59590 BP; 17614 A; 12031 C; 13575 G; 16118 T; 252 other;
 Query Match 100.0%; Score 18; DB 21; Length 59590;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

PI Preuss D, Copenhagen G, Keith K;
XX
XX WPI: 2000-587529/55.
XX
XX Recombinant DNA construct comprising a plant centromere, useful for
PT producing stably inherited microsome which can serve as vectors for
PT the construction of transgenic plant and animal cells
XX
XX
PS Claim 102; Page 529-549; 1449pp; English.
XX
XX The present invention relates to a recombinant DNA construct of a plant
CC (Arabidopsis thaliana) centromere. The constructs are useful for
CC producing stably inherited microsome which can serve as vectors for
CC the construction of transgenic plant and animal cells expressing
CC selected proteins such as hormones, enzymes, interleukins, clotting
CC factors, cytokines, antibodies, and growth factors.
XX
SQ Sequence 90336 BP; 26524 A; 18837 C; 18625 G; 26350 T; 0 other;
Query Match 100.0%; Score 18; DB 21; Length 90336;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 TGGCGCCGTTGCCAATTG 18
Db 84017 TGGCGCCGTTGCCAATTG 84000
RESULT 7
AAF22288/c
ID AAF22288 standard; DNA; 92584 BP.
XX
XX AAF22288;
AC
XX 20-MAR-2001 (first entry)
DT
XX BAC containing repeats from centromeres 1-4 #11.
DE
XX Centromere; microsome; vector; ds.
KW
XX Arabidopsis thaliana.
OS
XX WO200055325-A2.
PN
XX
XX 21-SEP-2000.
PD
XX
XX 17-MAR-2000; 2000WO-US07392.
PF
XX
XX 18-MAR-1999; 99US-0125219.
PR
XX 01-APR-1999; 99US-0127409.
PR
XX 18-MAY-1999; 99US-0134770.
PR
XX 13-SEP-1999; 99US-0153584.
PR
XX 17-SEP-1999; 99US-0154603.
PA
XX (UYCH-) UNIV CHICAGO.
PI
XX Preuss D, Copenhagen G, Keith K;
PI
XX WPI: 2000-587529/55.
DR
XX
XX Recombinant DNA construct comprising a plant centromere, useful for
PT producing stably inherited microsome which can serve as vectors for
PT the construction of transgenic plant and animal cells
XX
XX
PS Claim 102; Page 508-529; 1449pp; English.
XX
XX The present invention relates to a recombinant DNA construct of a plant
CC (Arabidopsis thaliana) centromere. The constructs are useful for
CC producing stably inherited microsome which can serve as vectors for
CC the construction of transgenic plant and animal cells expressing
CC selected proteins such as hormones, enzymes, interleukins, clotting
CC factors, cytokines, antibodies, and growth factors.
XX

SQ Sequence 92584 BP; 27840 A; 18113 C; 19835 G; 26796 T; 0 other;
Query Match 100.0%; Score 18; DB 21; Length 92584;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 TGGCGCCGTTGCCAATTG 18
Db 31015 TGGCGCCGTTGCCAATTG 30998
RESULT 8
AAF22302/c
ID AAF22302 standard; DNA; 94895 BP.
XX
XX AAF22302;
AC
XX 20-MAR-2001 (first entry)
DT
XX
XX BAC containing repeats from centromeres 1-4 #25.
DE
XX Centromere; microsome; vector; ds.
KW
XX Arabidopsis thaliana.
OS
XX WO200055325-A2.
PN
XX
XX 21-SEP-2000.
PD
XX
XX 17-MAR-2000; 2000WO-US07392.
PF
XX
XX 18-MAR-1999; 99US-0125219.
PR
XX 01-APR-1999; 99US-0127409.
PR
XX 18-MAY-1999; 99US-0134770.
PR
XX 13-SEP-1999; 99US-0153584.
PR
XX 17-SEP-1999; 99US-0154603.
PA
XX (UYCH-) UNIV CHICAGO.
PI
XX Preuss D, Copenhagen G, Keith K;
PI
XX WPI: 2000-587529/55.
DR
XX
XX Recombinant DNA construct comprising a plant centromere, useful for
PT producing stably inherited microsome which can serve as vectors for
PT the construction of transgenic plant and animal cells
XX
XX
PS Claim 102; Page 804-819; 1449pp; English.
XX
XX The present invention relates to a recombinant DNA construct of a plant
CC (Arabidopsis thaliana) centromere. The constructs are useful for
CC producing stably inherited microsome which can serve as vectors for
CC the construction of transgenic plant and animal cells expressing
CC selected proteins such as hormones, enzymes, interleukins, clotting
CC factors, cytokines, antibodies, and growth factors.
XX
SQ Sequence 94895 BP; 28943 A; 18101 C; 18466 G; 29385 T; 0 other;
Query Match 100.0%; Score 18; DB 21; Length 94895;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 TGGCGCCGTTGCCAATTG 18
Db 38282 TGGCGCCGTTGCCAATTG 38265
RESULT 9
AAF22282/c
ID AAF22282 standard; DNA; 95223 BP.
XX
XX AAF22282;
AC
XX

```

DT 20-MAR-2001 (first entry)
XX BAC containing repeats from centromeres 1-4 #5.
DE
XX Centromere; microsome; vector; ds.
XX
XX Arabidopsis thaliana.
OS
XX
XX WO200055325-A2.
PN
XX
XX 21-SEP-2000.
PD
XX
XX 17-MAR-2000; 2000WO-US07392.
PF
XX
XX 18-MAR-1999; 99US-0125219.
PR 01-APR-1999; 99US-0127409.
PR 18-MAY-1999; 99US-0134770.
PR 13-SEP-1999; 99US-0153584.
PR 17-SEP-1999; 99US-0154603.
XX
XX (UYCH-) UNIV CHICAGO.
XX
XX Preuss D, Copenhaver G, Keith K;
XX
XX WPI; 2000-587529/55.
DR
XX
XX Recombinant DNA construct comprising a plant centromere, useful for
PT producing stably inherited microsome which can serve as vectors for
PT the construction of transgenic plant and animal cells
XX
XX Claim 102; Page 364-385; 1449pp; English.
PS
XX
XX The present invention relates to a recombinant DNA construct of a plant
CC (Arabidopsis thaliana) centromere. The constructs are useful for
CC producing stably inherited microsome which can serve as vectors for
CC the construction of transgenic plant and animal cells expressing
CC selected proteins such as hormones, enzymes, interleukins, clotting
CC factors, cytokines, antibodies, and growth factors.
CC
XX
XX Sequence 95223 BP; 27974 A; 19452 C; 19355 G; 27745 T; 697 other;
SQ

```

Query Match 100.0%; Score 18; DB 21; Length 95223;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 TGGCGCCGTTGCCAATTG 18
    |||||||||||||||
Db 55836 TGGCGCCGTTGCCAATTG 55819

```

RESULT 10
 AAF22297/C
 ID AAF22297 standard; DNA; 96583 BP.
 XX
 AC AAF22297;
 XX
 DT 20-MAR-2001 (first entry)
 DE
 DE BAC containing repeats from centromeres 1-4 #20.
 XX
 XX Centromere; microsome; vector; ds.
 XX
 XX Arabidopsis thaliana.
 OS
 XX
 XX WO200055325-A2.
 PN
 XX
 XX 21-SEP-2000.
 PD
 XX
 XX 17-MAR-2000; 2000WO-US07392.
 PF
 XX
 XX 18-MAR-1999; 99US-0125219.
 PR 01-APR-1999; 99US-0127409.
 PR 18-MAY-1999; 99US-0134770.
 PR 13-SEP-1999; 99US-0153584.
 PR 17-SEP-1999; 99US-0154603.

```

PR 13-SEP-1999; 99US-0153584.
PR 17-SEP-1999; 99US-0154603.
XX
XX (UYCH-) UNIV CHICAGO.
XX
XX Preuss D, Copenhaver G, Keith K;
XX
XX WPI; 2000-587529/55.
DR
XX
XX Recombinant DNA construct comprising a plant centromere, useful for
PT producing stably inherited microsome which can serve as vectors for
PT the construction of transgenic plant and animal cells
XX
XX Claim 102; Page 716-738; 1449pp; English.
PS
XX
XX The present invention relates to a recombinant DNA construct of a plant
CC (Arabidopsis thaliana) centromere. The constructs are useful for
CC producing stably inherited microsome which can serve as vectors for
CC the construction of transgenic plant and animal cells expressing
CC selected proteins such as hormones, enzymes, interleukins, clotting
CC factors, cytokines, antibodies, and growth factors.
CC
XX
XX Sequence 96583 BP; 29910 A; 18944 C; 18341 G; 29388 T; 0 other;
SQ

```

Query Match 100.0%; Score 18; DB 21; Length 96583;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 TGGCGCCGTTGCCAATTG 18
    |||||||||||||||
Db 1458 TGGCGCCGTTGCCAATTG 1441

```

```

RESULT 11
AAF22290
ID AAF22290 standard; DNA; 96988 BP.
XX
AC AAF22290;
XX
XX 20-MAR-2001 (first entry)
DE
DE BAC containing repeats from centromeres 1-4 #13.
XX
XX Centromere; microsome; vector; ds.
XX
XX Arabidopsis thaliana.
OS
XX
XX WO200055325-A2.
PN
XX
XX 21-SEP-2000.
PD
XX
XX 17-MAR-2000; 2000WO-US07392.
PF
XX
XX 18-MAR-1999; 99US-0125219.
PR 01-APR-1999; 99US-0127409.
PR 18-MAY-1999; 99US-0134770.
PR 13-SEP-1999; 99US-0153584.
PR 17-SEP-1999; 99US-0154603.
XX
XX (UYCH-) UNIV CHICAGO.
XX
XX Preuss D, Copenhaver G, Keith K;
XX
XX WPI; 2000-587529/55.
DR
XX
XX Recombinant DNA construct comprising a plant centromere, useful for
PT producing stably inherited microsome which can serve as vectors for
PT the construction of transgenic plant and animal cells
XX
XX Claim 102; Page 550-572; 1449pp; English.
PS
XX
XX The present invention relates to a recombinant DNA construct of a plant
CC (Arabidopsis thaliana) centromere. The constructs are useful for

```

CC Producing stably inherited mitochondria which can serve as vectors for
CC the construction of transgenic plant and animal cells expressing
CC selected proteins such as hormones, enzymes, interleukins, clotting
CC factors, cytokines, antibodies, and growth factors.

Sequence 96988 BP; 29654 A; 18289 C; 18523 G; 30522 T; 0 other;

Query Match 100.0%; Score 18; DB 21; Length 96988;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGCGCGCTTGCACATG 18
DB 49874 TGGCGCGCTTGCACATG 49891

RESULT 12
AAE22298
ID AAE22298 standard; DNA; 109973 BP.

AC AAE22298;

DT 20-MAR-2001 (first entry)

DE BAC containing repeats from centromeres 1-4 #21.

DE Centromere; mitochondria; vector; ds.

OS Arabidopsis thaliana.

PN WO200055325-A2.

PD 21-SEP-2000.

PF 17-MAR-2000; 2000MO-US07392.

PR 18-MAR-1999; 99US-0125219.

PR 01-APR-1999; 99US-0127409.

PR 18-MAY-1999; 99US-0134770.

PR 13-SEP-1999; 99US-0153584.

PR 17-SEP-1999; 99US-0154603.

XX (UYCH-) UNIV CHICAGO.

XX Preuss D, Copenhaver G, Keith K;

XX WPI; 2000-587529/55.

XX Recombinant DNA construct comprising a plant centromere, useful for

XX producing stably inherited mitochondria which can serve as vectors for

XX the construction of transgenic plant and animal cells

XX Claim 102; Page 738-763; 1449pp; English.

XX The present invention relates to a recombinant DNA construct of a plant

XX (Arabidopsis thaliana) centromere. The constructs are useful for

XX producing stably inherited mitochondria which can serve as vectors for

XX the construction of transgenic plant and animal cells expressing

XX selected proteins such as hormones, enzymes, interleukins, clotting

XX factors, cytokines, antibodies, and growth factors.

XX Sequence 109973 BP; 31657 A; 23911 C; 22655 G; 31745 T; 5 other;

XX Query Match 100.0%; Score 18; DB 21; Length 109973;

XX Best Local Similarity 100.0%; Pred. No. 12;

XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGCGCGCTTGCACATG 18

DB 5453 TGGCGCGCTTGCACATG 5470

RESULT 13

AAE22286
ID AAE22286 standard; DNA; 134499 BP.

AC AAE22286;

DT 20-MAR-2001 (first entry)

DE BAC containing repeats from centromeres 1-4 #9.

DE Centromere; mitochondria; vector; ds.

OS Arabidopsis thaliana.

PN WO200055325-A2.

PD 21-SEP-2000.

PF 17-MAR-2000; 2000MO-US07392.

PR 18-MAR-1999; 99US-0125219.

PR 01-APR-1999; 99US-0127409.

PR 18-MAY-1999; 99US-0134770.

PR 13-SEP-1999; 99US-0153584.

PR 17-SEP-1999; 99US-0154603.

XX (UYCH-) UNIV CHICAGO.

XX Preuss D, Copenhaver G, Keith K;

XX WPI; 2000-587529/55.

XX Recombinant DNA construct comprising a plant centromere, useful for

XX producing stably inherited mitochondria which can serve as vectors for

XX the construction of transgenic plant and animal cells

XX Claim 102; Page 453-484; 1449pp; English.

XX The present invention relates to a recombinant DNA construct of a plant

XX (Arabidopsis thaliana) centromere. The constructs are useful for

XX producing stably inherited mitochondria which can serve as vectors for

XX the construction of transgenic plant and animal cells expressing

XX selected proteins such as hormones, enzymes, interleukins, clotting

XX factors, cytokines, antibodies, and growth factors.

XX Sequence 134499 BP; 41565 A; 25130 C; 25225 G; 42577 T; 2 other;

XX Query Match 100.0%; Score 18; DB 21; Length 134499;

XX Best Local Similarity 100.0%; Pred. No. 12;

XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGCGCGCTTGCACATG 18

DB 89656 TGGCGCGCTTGCACATG 89673

RESULT 14

AAE22303/c

ID AAE22303 standard; DNA; 611590 BP.

AC AAE22303;

DT 20-MAR-2001 (first entry)

DE Arabidopsis thaliana chromosome 2 centromere.

DE Centromere; mitochondria; vector; ds.

OS Arabidopsis thaliana.

PN WO200055325-A2.

PD 21-SEP-2000.

XX

PF 17-MAR-2000; 2000WO-US07392.
XX
PR 18-MAR-1999; 99US-0125219.
PR 01-APR-1999; 99US-0127409.
PR 18-MAY-1999; 99US-0134770.
PR 13-SEP-1999; 99US-0153584.
PR 17-SEP-1999; 99US-0154603.
XX
PA (UYCH-) UNIV CHICAGO.
XX
PI Preuss D, Copenhaver G, Keith K;
XX
DR WPI; 2000-587529/55.
XX
PT Recombinant DNA construct comprising a plant centromere, useful for
PT producing stably inherited microchromosomes which can serve as vectors for
PT the construction of transgenic plant and animal cells
XX
PS Claim 45; Page 820-959; 1449pp; English.
XX
CC The present invention relates to a recombinant DNA construct of a plant
CC (Arabidopsis thaliana) centromere. The constructs are useful for
CC producing stably inherited microchromosomes which can serve as vectors for
CC the construction of transgenic plant and animal cells expressing
CC selected proteins such as hormones, enzymes, interleukins, clotting
CC factors, cytokines, antibodies, and growth factors.
XX
SQ Sequence 611590 BP; 181893 A; 124460 C; 120254 G; 184983 T; 0 other;
OY
Query Match 100.0%; Score 18; DB 21; Length 611590;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 TGGCGCCGTTGCCAATTG 18
|||||
94448 TGGCGCCGTTGCCAATTG 94431
DB
RESULT 15
AAF22305
ID AAF22305 standard; DNA; 1082138 BP.
XX
AC AAF22305;
XX
DT 20-MAR-2001 (first entry)
XX
DE Arabidopsis thaliana chromosome 4 centromere.
XX
KM Centromere; microsome; vector; ds.
XX
OS Arabidopsis thaliana.
XX
PN WO200055325-A2.
XX
PD 21-SEP-2000.
XX
PE 17-MAR-2000; 2000WO-US07392.
XX
PR 18-MAR-1999; 99US-0125219.
PR 01-APR-1999; 99US-0127409.
PR 18-MAY-1999; 99US-0134770.
PR 13-SEP-1999; 99US-0153584.
PR 17-SEP-1999; 99US-0154603.
XX
PA (UYCH-) UNIV CHICAGO.
XX
PI Preuss D, Copenhaver G, Keith K;
XX
DR WPI; 2000-587529/55.
XX
PT Recombinant DNA construct comprising a plant centromere, useful for
PT producing stably inherited microchromosomes which can serve as vectors for
PT the construction of transgenic plant and animal cells

XX
PS Claim 68; Page 977-1388; 1449pp; English.
XX
CC The present invention relates to a recombinant DNA construct of a plant
CC (Arabidopsis thaliana) centromere. The constructs are useful for
CC producing stably inherited microchromosomes which can serve as vectors for
CC the construction of transgenic plant and animal cells expressing
CC selected proteins such as hormones, enzymes, interleukins, clotting
CC factors, cytokines, antibodies, and growth factors.
XX
SQ Sequence 1082138 BP; 348775 A; 194404 C; 195515 G; 343444 T; 0 other;
OY
Query Match 100.0%; Score 18; DB 21; Length 1082138;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 TGGCGCCGTTGCCAATTG 18
|||||
1043856 TGGCGCCGTTGCCAATTG 1043873
DB

Search completed: June 20, 2003, 21:58:19
Job time : 13.3019 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2003, 21:44:01 ; Search time 2.17925 Seconds
(without alignments)
2533.070 Million cell updates/sec

Title: US-09-965-553-1
Perfect score: 18
Sequence: 1 tggcgcgcgttgcattg 18

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: Issued_Patents_NA:*
2: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/Backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	18	100.0	18	US-09-322-478-1
2	18	100.0	150	US-09-322-478-36
3	15.4	85.6	1713	US-08-386-727-5
4	15.4	85.6	1713	US-08-600-452A-5
5	14.8	82.2	1272	US-09-053-702-1
6	14.8	82.2	3097	US-09-282-147-38
7	14.4	80.0	962	US-09-072-596-310
8	14.4	80.0	9515	US-08-920-812-13
9	14.4	80.0	9515	US-08-920-827-13
10	14.4	80.0	9515	US-08-921-177-13
11	14.4	80.0	9515	US-08-362-577C-13
12	14.4	80.0	9515	US-08-920-828-13
13	14.4	80.0	4403765	US-09-103-840A-2
14	14.4	80.0	4403765	US-09-103-840A-2
15	14.4	80.0	4411529	US-09-103-840A-1
16	14.4	80.0	4411529	US-09-103-840A-1
17	14.4	80.0	1386	US-08-247-475-16
18	14.4	77.8	1386	US-08-479-650-16
19	14.4	77.8	1386	US-08-191-866D-20
20	14.4	77.8	1386	US-08-674-169-16
21	14.4	77.8	1386	US-08-185-949B-20
22	13.8	76.7	897	US-09-556-877-120
23	13.8	76.7	897	US-09-556-877-130
24	13.8	76.7	897	US-09-556-877-132
25	13.8	76.7	897	US-09-556-877-132
26	13.8	76.7	897	US-09-620-412C-120
27	13.8	76.7	897	US-09-620-412C-130
				US-09-620-412C-132

ALIGNMENTS

c	28	13.8	76.7	1659	1	US-08-548-509-4	Sequence 4, Appl
c	29	13.8	76.7	3538	4	US-09-336-447A-10	Sequence 10, Appl
c	30	13.8	76.7	4588	4	US-09-453-702B-26	Sequence 26, Appl
c	31	13.8	76.7	10684	4	US-09-221-017B-401	Sequence 401, Appl
c	32	13.8	76.7	24417	2	US-08-846-762-1	Sequence 1, Appl
c	33	13.4	74.4	1596	2	US-08-531-601-2	Sequence 2, Appl
c	34	13.4	74.4	1596	2	US-08-859-032-2	Sequence 2, Appl
c	35	13.4	74.4	3390	4	US-09-550-338-1	Sequence 1, Appl
c	36	13.4	74.4	3825	1	US-08-737-597-1	Sequence 1, Appl
c	37	13.4	74.4	3825	1	US-08-737-597-2	Sequence 2, Appl
c	38	13.4	74.4	6312	1	US-08-531-601-3	Sequence 3, Appl
c	39	13.4	74.4	6312	2	US-08-859-032-3	Sequence 3, Appl
c	40	13.4	74.4	68750	3	US-09-335-409-1	Sequence 1, Appl
c	41	13.4	74.4	68750	4	US-09-568-102-1	Sequence 1, Appl
c	42	13.4	74.4	68750	4	US-09-567-969-1	Sequence 1, Appl
c	43	13.4	74.4	68750	4	US-09-568-480-1	Sequence 1, Appl
c	44	13.4	74.4	68750	4	US-09-568-486-1	Sequence 1, Appl
c	45	13.4	74.4	68750	4	US-09-568-472-1	Sequence 1, Appl

```

RESULT 1
US-09-322-478-1
; Sequence 1, Application US/09322478
; Patent No. 6331662
; GENERAL INFORMATION:
; APPLICANT: Wright, David A.
; APPLICANT: Voytas, Daniel F.
; TITLE OF INVENTION: Plant Retroelements and Methods Related Thereto
; FILE REFERENCE: P-1065 ISURF Plant Retroelement
; CURRENT APPLICATION NUMBER: US/09/322,478
; EARLIER FILING DATE: 1999-05-28
; EARLIER APPLICATION NUMBER: 60/087125
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 18
; TYPE: DNA
; ORGANISM: Glycine max
US-09-322-478-1

Query Match      100.0%; Score 18; DB 4; Length 18;
Best Local Similarity 100.0%; Pred. No. 0.74;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TGGCGCGGTGCCAATTG 18
DB      1 TGGCGCGGTGCCAATTG 18

RESULT 2
US-09-322-478-36
; Sequence 36, Application US/09322478
; Patent No. 6331662
; GENERAL INFORMATION:
; APPLICANT: Wright, David A.
; APPLICANT: Voytas, Daniel F.
; TITLE OF INVENTION: Plant Retroelements and Methods Related Thereto
; FILE REFERENCE: P-1065 ISURF Plant Retroelement
; CURRENT APPLICATION NUMBER: US/09/322,478
; EARLIER FILING DATE: 1999-05-28
; EARLIER APPLICATION NUMBER: 60/087125
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 36
; LENGTH: 150
; TYPE: DNA
; ORGANISM: Glycine max
US-09-322-478-36

```

Query Match 100.0%; Score 18; DB 4; Length 150;
Best Local Similarity 100.0%; Pred. No. 0.94;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGCGCGGTGCCAATTG 18
|||||
Db 66 TGGCGCGGTGCCAATTG 83

RESULT 3

US-08-386-727-5/c
; Sequence 5, Application US/08386727
; Patent No. 5792647
; GENERAL INFORMATION:
; APPLICANT: ROSEMAN, SAUL
; APPLICANT: BASSLER, BONNIE
; APPLICANT: KEYHANT, NEMAT O.
; APPLICANT: CHITLARD, EDITH
; APPLICANT: ROME, CHRIS
; APPLICANT: YU, CHARLES
; TITLE OF INVENTION: BACTERIAL CATABOLISM OF CHITIN
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: CUSHMAN, DARBY & CUSHMAN
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/386,727
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: HOBBS, ANN S.
; REGISTRATION NUMBER: 36,830
; REFERENCE/DOCKET NUMBER: 4130/206916
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-861-3000
; TELEFAX: 202-822-0944
; TELEX: 6714627 CUSH

; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1713 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-386-727-5

Query Match 85.6%; Score 15.4; DB 1; Length 1713;
Best Local Similarity 94.1%; Pred. No. 28;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GCGCGCGTGCATTG 18
|||||
Db 591 GCGCGCGTGCATTG 575

RESULT 4

US-08-600-452A-5/c
; Sequence 5, Application US/08600452A
; Patent No. 5985644
; GENERAL INFORMATION:
; APPLICANT: ROSEMAN, SAUL
; APPLICANT: BASSLER, BONNIE
; APPLICANT: KEYHANT, NEMAT O.

APPLICANT: CHITLARD, EDITH
; APPLICANT: ROME, CHRIS
; APPLICANT: YU, CHARLES
; TITLE OF INVENTION: BACTERIAL CATABOLISM OF CHITIN
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: FISH & RICHARDSON P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/600,452A
; FILING DATE: 13-FEB-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Hallie, Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 07662/005001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 678-5070
; TELEFAX: (619) 678-5099
; TELEX:

; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1713 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-600-452A-5

Query Match 85.6%; Score 15.4; DB 2; Length 1713;
Best Local Similarity 94.1%; Pred. No. 28;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GCGCGCGTGCATTG 18
|||||
Db 591 GCGCGCGTGCATTG 575

RESULT 5

US-09-053-702-1
; Sequence 1, Application US/09053702
; Patent No. 6229069
; GENERAL INFORMATION:
; APPLICANT: YAMADA, Shigehiro
; TITLE OF INVENTION: METHOD FOR CONTROLLING WATER CONTENT OF PLANT
; FILE REFERENCE: 230-122P
; CURRENT APPLICATION NUMBER: US/09/053,702
; CURRENT FILING DATE: 1998-04-02
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO: 1
; LENGTH: 1272
; TYPE: DNA
; ORGANISM: Mesembryanthemum crystallinum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (225)..(1070)
; US-09-053-702-1

Query Match 82.2%; Score 14.8; DB 4; Length 1272;
Best Local Similarity 88.9%; Pred. No. 56;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGGCGCGGTGCCAATTG 18

DB 841 TGGCTCCATTGCCAATTG 858

RESULT 6

US-09-282-147-38/C
Sequence 38, Application US/09282147
Patent No. 6274147
GENERAL INFORMATION:
APPLICANT: YAKHARIA, VIKRAM
APPLICANT: YAO, KUN
TITLE OF INVENTION: METHOD FOR GENERATING NONPATHOGENIC, INFECTIOUS
TITLE OF INVENTION: PANCREATIC NECROSIS VIRUS (IPNV) FROM SYNTHETIC RNA
FILE REFERENCE: 8288-9023
CURRENT APPLICATION NUMBER: US/09/282,147
CURRENT FILING DATE: 1999-03-31
EARLIER APPLICATION NUMBER: US/60/080,278
EARLIER FILING DATE: 1998-03-31
EARLIER APPLICATION NUMBER: PCT/US97/12955
EARLIER FILING DATE: 1998-03-31
NUMBER OF SEQ ID NOS: 51
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 38
LENGTH: 3097
TYPE: DNA
ORGANISM: Infectious pancreatic necrosis virus
US-09-282-147-38

Query Match 82.2%; Score 14.8; DB 4; Length 3097;
Best Local Similarity 88.9%; Pred. No. 62;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGGCGCCGTTGCCAATTG 18
DB 1091 TGGCGACGTTGCCGATTG 1074

RESULT 7

US-09-072-596-310/C
Sequence 310, Application US/09072596
Patent No. 6458366
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Neto, Antonia
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.
APPLICANT: Lodes, Michael J.
APPLICANT: Hendrickson, Ronald C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
NUMBER OF SEQUENCES: 350
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/072,596
FILING DATE: 05-MAY-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392

REFERENCE/DOCKET NUMBER: 210121.417C9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 310:
SEQUENCE CHARACTERISTICS:
LENGTH: 962 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
US-09-072-596-310

Query Match 80.0%; Score 14.4; DB 4; Length 962;
Best Local Similarity 93.8%; Pred. No. 88;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGGCGCCGTTGCCAAT 16
DB 142 TGGCGCCGTTGCCGAT 127

RESULT 8

US-08-920-812-13/C
Sequence 13, Application US/08920812
Patent No. 5763188
GENERAL INFORMATION:
APPLICANT: Ohno, Tsuneya
APPLICANT: Matsuhisa, Akio
APPLICANT: Uehara, Hirotsugu
APPLICANT: Eda, Soji
TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/920,812
FILING DATE: 29-AUG-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/362,577
FILING DATE: 27-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Rin-Laures, Li-Hsien
REGISTRATION NUMBER: 33,547
REFERENCE/DOCKET NUMBER: 19036/32420
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
TELEX: 25-3856

INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 9515 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
ORIGINAL SOURCE:
ORGANISM: Pseudomonas aeruginosa
STRAIN: Clinical isolate P2-2
US-08-920-812-13

Query Match 80.0%; Score 14.4; DB 1; Length 9515;

Best Local Similarity 93.8%; Pred. No. 1.1e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 GCGCGCTGGCAATG 18
|||||
Db 515 GCGCGCTGGCAATCG 500

RESULT 9

US-08-920-827-13/C
Sequence 13, Application US/08920827
Patent No. 5770375

GENERAL INFORMATION:

APPLICANT: Ohno, Tsuneya
APPLICANT: Matsuhisa, Akio
APPLICANT: Uehara, Hirotsugu
APPLICANT: Eda, Soji
TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402

COMPUTER READABLE FORM:

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/920,827
FILING DATE: 29-AUG-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/362,577
FILING DATE: 27-MAR-1995

ATTORNEY/AGENT INFORMATION:

NAME: Rin-Laures, Li-Hsien
REGISTRATION NUMBER: 33,547
REFERENCE/DOCKET NUMBER: 19036/32420
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448

TELECOMMUNICATION INFORMATION:

TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:

LENGTH: 9515 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA

ORIGINAL SOURCE:

ORGANISM: Pseudomonas aeruginosa
STRAIN: Clinical Isolate P2-2
US-08-920-827-13

Query Match 80.0%; Score 14.4; DB 1; Length 9515;
Best Local Similarity 93.8%; Pred. No. 1.1e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 GCGCGCTGGCAATG 18
|||||
Db 515 GCGCGCTGGCAATCG 500

RESULT 10

US-08-921-177-13/C
Sequence 13, Application US/08921177
Patent No. 5798211
GENERAL INFORMATION:
APPLICANT: Ohno, Tsuneya

APPLICANT: Matsuhisa, Akio
APPLICANT: Uehara, Hirotsugu
APPLICANT: Eda, Soji
TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402

COMPUTER READABLE FORM:

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/921,177
FILING DATE: 29-AUG-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/362,577
FILING DATE: 27-MAR-1995

ATTORNEY/AGENT INFORMATION:

NAME: Rin-Laures, Li-Hsien
REGISTRATION NUMBER: 33,547
REFERENCE/DOCKET NUMBER: 19036/32420
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448

INFORMATION FOR SEQ ID NO: 13:

SEQUENCE CHARACTERISTICS:
LENGTH: 9515 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA

ORIGINAL SOURCE:

ORGANISM: Pseudomonas aeruginosa
STRAIN: Clinical Isolate P2-2
US-08-921-177-13

Query Match 80.0%; Score 14.4; DB 1; Length 9515;
Best Local Similarity 93.8%; Pred. No. 1.1e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 GCGCGCTGGCAATG 18
|||||
Db 515 GCGCGCTGGCAATCG 500

RESULT 11

US-08-362-577C-13/C
Sequence 13, Application US/08362577C
Patent No. 5807673

GENERAL INFORMATION:

APPLICANT: Ohno, Tsuneya
APPLICANT: Matsuhisa, Akio
APPLICANT: Uehara, Hirotsugu
APPLICANT: Eda, Soji
TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402

COMPUTER READABLE FORM:

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/362,577C
FILING DATE: 27-MAR-1995
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Rin-Laures, Li-Hsien
REGISTRATION NUMBER: 33,547
REFERENCE/DOCKET NUMBER: 19036/32420
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ. ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 9515 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
ORIGINAL SOURCE:
ORGANISM: Pseudomonas aeruginosa
STRAIN: Clinical Isolate P2-2
US-08-362-577C-13

Query Match 80.0%; Score 14.4; DB 1; Length 9515;
Best Local Similarity 93.8%; Pred. No. 1.1e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 GCGCCGTGCGCATTTG 18
|||||
Db 515 GCGCCGTGCGCATTCG 500

RESULT 12
US-08-920-828-13/C
Sequence 13, Application US/08920828
Patent No. 5853998
GENERAL INFORMATION:
APPLICANT: Ohno, Tsuneya
APPLICANT: Matsuhisa, Akio
APPLICANT: Uehara, Hirotsugu
APPLICANT: Edo, Soji
TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/920,828
FILING DATE: 29-AUG-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/362,577
FILING DATE: 27-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Rin-Laures, Li-Hsien
REGISTRATION NUMBER: 33,547
REFERENCE/DOCKET NUMBER: 19036/32420
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448

TELEX: 25-3856
INFORMATION FOR SEQ. ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 9515 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
ORIGINAL SOURCE:
ORGANISM: Pseudomonas aeruginosa
STRAIN: Clinical Isolate P2-2
US-08-920-828-13

Query Match 80.0%; Score 14.4; DB 2; Length 9515;
Best Local Similarity 93.8%; Pred. No. 1.1e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 GCGCCGTGCGCATTTG 18
|||||
Db 515 GCGCCGTGCGCATTCG 500

RESULT 13
US-09-103-840A-2
Sequence 2, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 2
LENGTH: 4403765
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
FEATURE:
OTHER INFORMATION: CDC 1551
OTHER INFORMATION: "n" bases at various positions throughout the sequence
US-09-103-840A-2

Query Match 80.0%; Score 14.4; DB 4; Length 4403765;
Best Local Similarity 93.8%; Pred. No. 67;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TGGCGCCGTGCGCAAT 16
|||||
Db 1572527 TGGCGCCGTGCGCAAT 1572542

RESULT 14
US-09-103-840A-2/C
Sequence 2, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: Patentin Ver. 2.1

SEQ ID NO 2
LENGTH: 4403765
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
FEATURE:
OTHER INFORMATION: CDC 1551
OTHER INFORMATION: "n" bases at various positions throughout the sequence
OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 80.0%; Score 14.4; DB 4; Length 4403765;
Best Local Similarity 93.8%; Pred. No. 67;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGGCGCGGTGCCAAT 16
|||
Db 841864 TGGCGCGGTGCCGAT 841849

RESULT 15
US-09-103-840A-1
Sequence 1, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 4411529
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
OTHER INFORMATION: H37RV
US-09-103-840A-1

Query Match 80.0%; Score 14.4; DB 4; Length 4411529;
Best Local Similarity 93.8%; Pred. No. 67;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGGCGCGGTGCCAAT 16
|||
Db 1572688 TGGCGCGGTGCCAAT 1572703

Search completed: June 20, 2003, 23:25:32
Job time: 12.1792 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2003, 21:58:41 ; Search time 6.65094 Seconds
(without alignments)
3971.420 Million cell updates/sec

Title: US-09-965-553-1

Perfect score: 18

Sequence: 1 tggcgccgttgccattg 18

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1042519 seqs, 733713590 residues

Total number of hits satisfying chosen parameters: 2085038

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA.*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	18	100.0	18	US-09-965-553-1	Sequence 1, Appl
2	18	100.0	150	US-09-965-553-36	Sequence 36, Appl
3	15	83.3	412	US-10-101-464A-479	Sequence 479, Appl
4	14.8	82.2	266	US-09-878-574-15377	Sequence 15377, A
5	14.8	82.2	573	US-09-974-300-4705	Sequence 4705, A
6	14.8	82.2	861	US-09-938-842A-807	Sequence 807, Appl
7	14.8	82.2	1089	US-09-974-300-764	Sequence 764, Appl
8	14.8	82.2	2155	US-10-002-050-11	Sequence 11, Appl
9	14.8	82.2	2155	US-10-002-304-11	Sequence 11, Appl
10	14.8	82.2	2155	US-10-003-152-11	Sequence 11, Appl
11	14.8	82.2	2156	US-10-002-050-21	Sequence 21, Appl
12	14.8	82.2	2156	US-10-002-304-21	Sequence 21, Appl
13	14.8	82.2	2156	US-10-003-152-21	Sequence 21, Appl
14	14.8	82.2	2284	US-10-002-050-13	Sequence 13, Appl
15	14.8	82.2	2284	US-10-002-304-13	Sequence 13, Appl
16	14.8	82.2	2284	US-10-003-152-13	Sequence 13, Appl
17	14.8	82.2	3097	US-10-216-981A-25	Sequence 25, Appl
18	14.8	82.2	3293	US-10-149-819-25	Sequence 25, Appl
19	14.8	82.2	8095	US-09-989-920-73	Sequence 73, Appl

C 20	14.4	80.0	943	10	US-09-822-830A-533	Sequence 533, App
C 21	14.4	80.0	1854	9	US-09-894-844-103	Sequence 103, App
C 22	14.4	80.0	1863	10	US-09-815-242-9985	Sequence 9985, App
C 23	14.4	77.8	1392	10	US-09-815-242-9820	Sequence 9820, App
C 24	14.4	77.8	3268	9	US-09-927-827-24	Sequence 24, Appl
C 25	13.8	76.7	174	9	US-10-122-822-21	Sequence 21, Appl
C 26	13.8	76.7	365	10	US-09-880-107-357	Sequence 357, App
C 27	13.8	76.7	476	10	US-09-974-300-5603	Sequence 5603, App
C 28	13.8	76.7	538	9	US-09-925-299-369	Sequence 369, App
C 29	13.8	76.7	538	10	US-09-925-299-369	Sequence 369, App
C 30	13.8	76.7	576	10	US-09-974-300-5436	Sequence 5436, App
C 31	13.8	76.7	897	10	US-09-841-132-120	Sequence 120, App
C 32	13.8	76.7	897	10	US-09-841-132-130	Sequence 130, App
C 33	13.8	76.7	897	10	US-09-841-132-132	Sequence 132, App
C 34	13.8	76.7	1078	10	US-09-974-300-2700	Sequence 2700, App
C 35	13.8	76.7	1320	10	US-09-815-242-9819	Sequence 9819, App
C 36	13.8	76.7	1405	9	US-10-122-822-6	Sequence 6, Appl
C 37	13.8	76.7	1627	12	US-10-062-254-237	Sequence 237, App
C 38	13.8	76.7	2940	10	US-09-801-368-283	Sequence 283, App
C 39	13.8	76.7	3538	9	US-09-952-267-10	Sequence 10, Appl
C 40	13.8	76.7	4588	9	US-10-114-170-26	Sequence 26, Appl
C 41	13.8	76.7	6373	9	US-10-000-512-1	Sequence 1, Appl
C 42	13.8	76.7	6378	9	US-10-037-270-332	Sequence 332, App
C 43	13.8	76.7	13715	7	US-08-781-986A-195	Sequence 195, App
C 44	13.8	76.7	465237	10	US-09-933-267A-1	Sequence 1, Appl
C 45	13.4	74.4	277	10	US-09-294-093B-2256	Sequence 2256, App

ALIGNMENTS

RESULT 1
US-09-965-553-1
; Sequence 1, Application US/09965553
; Patent No. US20020112259A1
; GENERAL INFORMATION:
; APPLICANT: Wright, David A.
; TITLE OF INVENTION: Plant Retroelements and Methods Related Thereto
; FILE REFERENCE: P-1065 ISURF Plant Retroelement
; CURRENT APPLICATION NUMBER: US/09/965,553
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 09/322,478
; PRIOR FILING DATE: 1999-05-28
; PRIOR APPLICATION NUMBER: 60/087125
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 18
; TYPE: DNA
; ORGANISM: Glycine max
US-09-965-553-1
Query Match 100.0%; Score 18; DB 10; Length 18;
Best Local Similarity 100.0%; Pred No. 1.8;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DB 1 TGGCGCCGTTGCCAATTG 18
OY ||||||||||||||||
1 TGGCGCCGTTGCCAATTG 18
RESULT 2
US-09-965-553-36
; Sequence 36, Application US/09965553
; Patent No. US20020112259A1
; GENERAL INFORMATION:
; APPLICANT: Wright, David A.
; APPLICANT: Voytas, Daniel F.
; TITLE OF INVENTION: Plant Retroelements and Methods Related Thereto
; FILE REFERENCE: P-1065 ISURF Plant Retroelement
; CURRENT APPLICATION NUMBER: US/09/965,553

;; CURRENT FILING DATE: 2001-09-27
;; PRIOR APPLICATION NUMBER: 09/322,478
;; PRIOR FILING DATE: 1999-05-28
;; PRIOR APPLICATION NUMBER: 60/087125
;; PRIOR FILING DATE: 1998-05-29
;; NUMBER OF SEQ ID NOS: 41
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 36
;; LENGTH: 150
;; TYPE: DNA
;; ORGANISM: Glycine max
;; US-09-965-553-36

Query Match 100.0%; Score 18; DB 10; Length 150;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGCGCCGTTGCCAATTG 18
|||||
Db 66 TGGCGCCGTTGCCAATTG 83

RESULT 3
US-10-101-464A-479/C
;; Sequence 479, Application US/10101464A
;; Publication No. US20030046728A1
;; GENERAL INFORMATION:
;; APPLICANT: Strabala, Timothy
;; APPLICANT: Nieuwenhuizen, Nicolaas
;; APPLICANT: Higgins, Colleen M.
;; TITLE OF INVENTION: Compositions isolated from plant cells
;; TITLE OF INVENTION: and their use in the modification of plant cell signaling
;; FILE REFERENCE: 11000.1020c2
;; CURRENT APPLICATION NUMBER: US/10/101,464A
;; CURRENT FILING DATE: 2002-03-18
;; PRIOR APPLICATION NUMBER: 09/704,302
;; PRIOR FILING DATE: 2000-11-01
;; PRIOR APPLICATION NUMBER: 09/228,986
;; PRIOR FILING DATE: 1999-01-12
;; PRIOR APPLICATION NUMBER: 60/162,866
;; PRIOR FILING DATE: 1999-11-01
;; PRIOR APPLICATION NUMBER: PCT/US00/00724
;; PRIOR FILING DATE: 2000-01-11
;; NUMBER OF SEQ ID NOS: 989
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 479
;; LENGTH: 412
;; TYPE: DNA
;; ORGANISM: Pinus radiata
;; US-10-101-464A-479

Query Match 83.3%; Score 15; DB 9; Length 412;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GCGCGCTTGCCATT 17
|||||
Db 32 GCGCGCTTGCCATT 18

RESULT 4
US-09-878-574-15377
;; Sequence 15377, Application US/09878574
;; Patent No. US20020110548A1
;; GENERAL INFORMATION:
;; APPLICANT: Byrum, Joseph R.
;; APPLICANT: La Rosa, Thomas J.
;; APPLICANT: Thompson, Michael D.
;; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
;; TITLE OF INVENTION: Plants
;; FILE REFERENCE: 38-21115401B
;; CURRENT APPLICATION NUMBER: US/09/878,574
;; CURRENT FILING DATE: 2001-12-21

;; PRIOR APPLICATION NUMBER: 09/333,535
;; PRIOR FILING DATE: 1999-06-14
;; NUMBER OF SEQ ID NOS: 15775
;; SEQ ID NO 15377
;; LENGTH: 266
;; TYPE: DNA
;; ORGANISM: Glycine max
;; OTHER INFORMATION: Clone ID: 701070032H2
;; US-09-878-574-15377

Query Match 82.2%; Score 14.8; DB 10; Length 266;
Best Local Similarity 88.9%; Pred. No. 1.3e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGGCGCCGTTGCCAATTG 18
|||||
Db 6 TGGCACCCTTGCCATTG 23

RESULT 5
US-09-974-300-4705
;; Sequence 4705, Application US/09974300
;; Patent No. US20020146721A1
;; GENERAL INFORMATION:
;; APPLICANT: Berka, Randy M.
;; APPLICANT: Clausen, Ib Groth
;; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
;; TITLE OF INVENTION: Expression
;; FILE REFERENCE: 10085 500-US
;; CURRENT APPLICATION NUMBER: US/09/974,300
;; CURRENT FILING DATE: 2001-10-05
;; PRIOR APPLICATION NUMBER: 09/680,598
;; PRIOR FILING DATE: 2000-10-06
;; PRIOR APPLICATION NUMBER: 60/279,526
;; PRIOR FILING DATE: 2001-03-27
;; NUMBER OF SEQ ID NOS: 8481
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 4705
;; LENGTH: 573
;; TYPE: DNA
;; ORGANISM: Bacillus clausii
;; US-09-974-300-4705

Query Match 82.2%; Score 14.8; DB 10; Length 573;
Best Local Similarity 88.9%; Pred. No. 1.4e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGGCGCCGTTGCCAATTG 18
|||||
Db 287 TGACGCCGTTGCCAATTG 304

RESULT 6
US-09-938-842A-807
;; Sequence 807, Application US/09938842A
;; Patent No. US20020160378A1
;; GENERAL INFORMATION:
;; APPLICANT: Harper, Jeff
;; APPLICANT: Kreps, Joel
;; APPLICANT: Wang, Xun
;; APPLICANT: Zhu, Tong
;; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
;; TITLE OF INVENTION: SAME, AND METHODS OF USE
;; FILE REFERENCE: SCRIPI300-3
;; CURRENT APPLICATION NUMBER: US/09/938,842A
;; CURRENT FILING DATE: 2001-08-24
;; PRIOR APPLICATION NUMBER: US 60/227,866
;; PRIOR FILING DATE: 2000-08-24
;; PRIOR APPLICATION NUMBER: US 60/264,647
;; PRIOR FILING DATE: 2001-01-16
;; PRIOR APPLICATION NUMBER: US 60/300,111
;; PRIOR FILING DATE: 2001-06-22
;; NUMBER OF SEQ ID NOS: 5379

SEQ ID NO 807
LENGTH: 861
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-938-842A-807

Query Match 82.2%; Score 14.8; DB 9; Length 861;
Best Local Similarity 88.9%; Pred. No. 1.4e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TGGCGCCGTTGCCAATTG 18
|||||
DB 605 TGGCTCCATTGCCAATTG 622

RESULT 7
US-09-974-300-764/C
Sequence 764, Application US/09974300
Patent No. US20020146721A1
GENERAL INFORMATION:
APPLICANT: Berka, Randy M.
APPLICANT: Clausen, Id Groth
TITLE OF INVENTION: Methods for Monitoring Multiple Gene
FILE REFERENCE: 10085, 500-US
CURRENT APPLICATION NUMBER: US/09/974, 300
CURRENT FILING DATE: 2001-10-05
PRIOR APPLICATION NUMBER: 09/680, 598
PRIOR FILING DATE: 2000-10-06
PRIOR APPLICATION NUMBER: 60/279, 526
PRIOR FILING DATE: 2001-03-27
NUMBER OF SEQ ID NOS: 8481
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 764
LENGTH: 1089
TYPE: DNA
ORGANISM: Bacillus licheniformis
US-09-974-300-764

Query Match 82.2%; Score 14.8; DB 10; Length 1089;
Best Local Similarity 88.9%; Pred. No. 1.5e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TGGCGCCGTTGCCAATTG 18
|||||
DB 1070 TGGCGCCGTTGCCAATTG 1053

RESULT 8
US-10-002-050-11/C
Sequence 11, Application US/10002050
Publication No. US20030032095A1
GENERAL INFORMATION:
APPLICANT: Shinkets, Richard
APPLICANT: Fernandes, Elma
APPLICANT: Vernet, Corine
APPLICANT: Yang, Meijia
APPLICANT: Boldog, Ferenc
TITLE OF INVENTION: No. US20030032095A1 Nucleic Acid Sequences Encoding Human Semaphorin
FILE REFERENCE: 15966-554 Cura-54 CON-S14
CURRENT APPLICATION NUMBER: US/10/002, 050
CURRENT FILING DATE: 2001-11-02
PRIOR APPLICATION NUMBER: 09/604, 286
PRIOR FILING DATE: 2000-06-22
PRIOR APPLICATION NUMBER: 60/140, 584
PRIOR FILING DATE: 1999-06-23
NUMBER OF SEQ ID NOS: 49
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 11
LENGTH: 2155
TYPE: DNA
ORGANISM: Homo sapiens

FEATURE:
NAME/KEY: CDS
LOCATION: (166)..(1935)
US-10-002-050-11

Query Match 82.2%; Score 14.8; DB 9; Length 2155;
Best Local Similarity 88.9%; Pred. No. 1.6e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TGGCGCCGTTGCCAATTG 18
|||||
DB 1298 TGGCGCCGATGCCAGTTG 1281

RESULT 9
US-10-002-304-11/C
Sequence 11, Application US/10002304
Publication No. US20030036185A1
GENERAL INFORMATION:
APPLICANT: Shinkets, Richard
APPLICANT: Fernandes, Elma
APPLICANT: Vernet, Corine
APPLICANT: Yang, Meijia
APPLICANT: Boldog, Ferenc
TITLE OF INVENTION: Polynucleotides and polypeptides encoded thereby
FILE REFERENCE: 15966-554 Cura-54 CON-S8
CURRENT APPLICATION NUMBER: US/10/002, 304
CURRENT FILING DATE: 2001-11-02
PRIOR APPLICATION NUMBER: 09/604, 286
PRIOR FILING DATE: 2000-06-22
PRIOR APPLICATION NUMBER: 60/140, 584
PRIOR FILING DATE: 1999-06-23
NUMBER OF SEQ ID NOS: 49
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 11
LENGTH: 2155
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (166)..(1935)
US-10-002-304-11

Query Match 82.2%; Score 14.8; DB 9; Length 2155;
Best Local Similarity 88.9%; Pred. No. 1.6e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TGGCGCCGTTGCCAATTG 18
|||||
DB 1298 TGGCGCCGATGCCAGTTG 1281

RESULT 10
US-10-003-152-11/C
Sequence 11, Application US/10003152
Patent No. US20020151494A1
GENERAL INFORMATION:
APPLICANT: Shinkets, Richard
APPLICANT: Fernandes, Elma
APPLICANT: Vernet, Corine
APPLICANT: Yang, Meijia
APPLICANT: Boldog, Ferenc
TITLE OF INVENTION: No. US20020151494A1 Amino Acid Sequences for Human Semaphorin
FILE REFERENCE: 15966-554 Cura-54 CON-S12
CURRENT APPLICATION NUMBER: US/10/003, 152
CURRENT FILING DATE: 2001-11-02
PRIOR APPLICATION NUMBER: 09/604, 286
PRIOR FILING DATE: 2000-06-22
PRIOR APPLICATION NUMBER: 60/140, 584
PRIOR FILING DATE: 1999-06-23
NUMBER OF SEQ ID NOS: 49

```
SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 2155
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (166)..(1935)
US-10-003-152-11

Query Match      82.2%; Score 14.8; DB 12; Length 2155;
Best Local Similarity 88.9%; Pred. No. 1.6e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TGGCGCCGTTGCCAATTG 18
        ||||||| ||||| |||
Db      1298 TGGCGCCGATGCCAGTTG 1281

RESULT 11
US-10-002-050-21/c
; Sequence 21, Application US/10002050
; Publication No. US20030032095A1
; GENERAL INFORMATION:
; APPLICANT: Shimkets, Richard
; APPLICANT: Fernandes, Elma
; APPLICANT: Vernet, Corline
; APPLICANT: Yang, Meijia
; APPLICANT: Boldog, Ferenc
; APPLICANT: Herrmann, John
; TITLE OF INVENTION: No. US20030032095A1 Nucleic Acid Sequences Encoding Human Semaph
; FILE REFERENCE: 15966-554 Cura-54 CON-S14
; CURRENT APPLICATION NUMBER: US/10/002,050
; PRIOR FILING DATE: 2001-11-02
; PRIOR APPLICATION NUMBER: 09/604,286
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 60/140,584
; PRIOR FILING DATE: 1999-06-23
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21
; LENGTH: 2156
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (166)..(2037)
; NAME/KEY: variation
; LOCATION: (1)..(2156)
; OTHER INFORMATION: N may be any nucleotide
US-10-002-050-21

Query Match      82.2%; Score 14.8; DB 9; Length 2156;
Best Local Similarity 88.9%; Pred. No. 1.6e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TGGCGCCGTTGCCAATTG 18
        ||||||| ||||| |||
Db      1298 TGGCGCCGATGCCAGTTG 1281

RESULT 12
US-10-002-304-21/c
; Sequence 21, Application US/10002304
; Publication No. US20030036185A1
; GENERAL INFORMATION:
; APPLICANT: Shimkets, Richard
; APPLICANT: Fernandes, Elma
; APPLICANT: Vernet, Corline
; APPLICANT: Yang, Meijia
; APPLICANT: Boldog, Ferenc
; APPLICANT: Herrmann, John
; TITLE OF INVENTION: Polynucleotides and polypeptides encoded thereby
```

```
FILE REFERENCE: 15966-554 Cura-54 CON-S8
; CURRENT APPLICATION NUMBER: US/10/002,304
; CURRENT FILING DATE: 2001-11-02
; PRIOR APPLICATION NUMBER: 09/604,286
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 60/140,584
; PRIOR FILING DATE: 1999-06-23
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21
; LENGTH: 2156
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (166)..(2037)
; NAME/KEY: variation
; LOCATION: (1)..(2156)
; OTHER INFORMATION: N may be any nucleotide
US-10-003-152-21

Query Match      82.2%; Score 14.8; DB 9; Length 2156;
Best Local Similarity 88.9%; Pred. No. 1.6e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TGGCGCCGTTGCCAATTG 18
        ||||||| ||||| |||
Db      1298 TGGCGCCGATGCCAGTTG 1281

RESULT 13
US-10-003-152-21/c
; Sequence 21, Application US/10003152
; Patent No. US20020151494A1
; GENERAL INFORMATION:
; APPLICANT: Shimkets, Richard
; APPLICANT: Fernandes, Elma
; APPLICANT: Vernet, Corline
; APPLICANT: Yang, Meijia
; APPLICANT: Boldog, Ferenc
; APPLICANT: Herrmann, John
; TITLE OF INVENTION: No. US20020151494A1 Amino Acid Sequences for Human Semaphorin
; FILE REFERENCE: 15966-554 Cura-54 CON-S12
; CURRENT APPLICATION NUMBER: US/10/003,152
; CURRENT FILING DATE: 2001-11-02
; PRIOR APPLICATION NUMBER: 09/604,286
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 60/140,584
; PRIOR FILING DATE: 1999-06-23
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21
; LENGTH: 2156
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (166)..(2037)
; NAME/KEY: variation
; LOCATION: (1)..(2156)
; OTHER INFORMATION: N may be any nucleotide
US-10-003-152-21

Query Match      82.2%; Score 14.8; DB 12; Length 2156;
Best Local Similarity 88.9%; Pred. No. 1.6e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TGGCGCCGTTGCCAATTG 18
        ||||||| ||||| |||
Db      1298 TGGCGCCGATGCCAGTTG 1281

RESULT 14
```


US-10-002-050-13/c
; Sequence 13, Application US/10002050
; Publication No. US20030032095A1
; GENERAL INFORMATION:
; APPLICANT: Shimkets, Richard
; APPLICANT: Fernandes, Elma
; APPLICANT: Vernet, Corine
; APPLICANT: Yang, Meijia
; APPLICANT: Boldog, Ferenc
; APPLICANT: Herrmann, John
; TITLE OF INVENTION: No. US20030032095A1 Nucleic Acid Sequences Encoding Human Semaphorin
; FILE REFERENCE: 15966-554 Cura-54 CON-S14
; CURRENT APPLICATION NUMBER: US/10/002,050
; CURRENT FILING DATE: 2001-11-02
; PRIOR APPLICATION NUMBER: 09/604,286
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 60/140,584
; PRIOR FILING DATE: 1999-06-23
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO: 13
; LENGTH: 2284
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (166)..(1953)
; NAME/KEY: variation
; LOCATION: (1)..(2284)
; OTHER INFORMATION: N may be any nucleotide
US-10-002-050-13

Query Match 82.2%; Score 14.8; DB 9; Length 2284;
Best Local Similarity 88.9%; Pred. No. 1.6e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGGCGCCGTTGCCAATTG 18
||||||| ||||| |||
DB 1298 TGGCGCCGATGCCAGTTG 1281

RESULT 15
US-10-002-304-13/c
; Sequence 13, Application US/10002304
; Publication No. US20030036185A1
; GENERAL INFORMATION:
; APPLICANT: Shimkets, Richard
; APPLICANT: Fernandes, Elma
; APPLICANT: Vernet, Corine
; APPLICANT: Yang, Meijia
; APPLICANT: Boldog, Ferenc
; APPLICANT: Herrmann, John
; TITLE OF INVENTION: Polynucleotides and polypeptides encoded thereby
; FILE REFERENCE: 15966-554 Cura-54 CON-S8
; CURRENT APPLICATION NUMBER: US/10/002,304
; CURRENT FILING DATE: 2001-11-02
; PRIOR APPLICATION NUMBER: 09/604,286
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 60/140,584
; PRIOR FILING DATE: 1999-06-23
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO: 13
; LENGTH: 2284
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (166)..(1953)
; NAME/KEY: variation
; LOCATION: (1)..(2284)
; OTHER INFORMATION: N may be any nucleotide
US-10-002-304-13

Query Match 82.2%; Score 14.8; DB 9; Length 2284;
Best Local Similarity 88.9%; Pred. No. 1.6e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGGCGCCGTTGCCAATTG 18
||||||| ||||| |||
DB 1298 TGGCGCCGATGCCAGTTG 1281

Search completed: June 21, 2003, 00:34:38
Job time : 8.65094 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2003, 19:31:16 ; Search time 2520.75 Seconds

(without alignments)
6927.165 Million cell updates/sec

Title: US-09-965-553-11

Perfect score: 600

Sequence: 1 ttggagcgtcgtcgcataata.....tccatagccacaagatctca 600

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	571.2	95.2	8573	8	AF186184
2	532	88.7	10128	8	AF186182
3	529.6	88.3	762	8	AF378068
4	526.4	87.7	762	8	AF378062
5	523.2	87.2	12886	8	AF186183
6	502.6	83.8	763	8	AF378063
7	430	71.7	762	8	AF378064
8	359	59.8	92281	8	AP004896
9	352.6	58.8	762	8	AF378072
10	351.6	58.6	762	8	AF378070
11	349.8	58.3	761	8	AF378073
12	348.8	58.1	761	8	AF378069
13	346.2	57.7	110729	2	AC123573
14	345.6	57.6	13637	8	AF186185
15	342.2	57.0	761	8	AF378071
16	340.4	56.7	762	8	AF378065
17	334	55.7	762	8	AF378067
18	332.6	55.4	761	8	AF378066
19	330.2	55.0	762	8	AF378049
20	327	54.5	762	8	AF378050
21	323.8	54.0	3147	8	AB007466
22	323.2	53.9	264461	2	AC130804
23	319.6	53.3	80930	8	AP004894
24	319	53.2	762	8	AF378051
25	319	53.1	2531	8	PSJ000639
26	315.8	52.6	12314	8	PSJ000640
27	313.6	52.3	760	8	AF378015
28	311.6	51.9	108720	8	AB046436
29	311.6	51.9	138940	8	AC063973
30	308.8	51.5	87814	8	AB046426
31	308.8	51.5	52616	2	AC090029
32	308.4	51.4	762	8	AF378057
33	308	51.3	60482	2	AC090030
34	307.8	51.3	18813	8	TCE19
35	307.8	51.3	35313	8	AB028613
36	307.8	51.3	38519	6	AX059548
37	307.8	51.3	48422	6	AX059509
38	307.8	51.3	94805	8	F2112
39	307.8	51.3	196766	8	ATCHR1Y16
40	307.2	51.2	85962	8	AB046431
41	307.2	51.2	95463	8	F1809
42	306.2	51.0	762	8	AF378079
43	306.2	51.0	108827	8	ATF7P3
44	306.2	51.0	108720	8	AB046436
45	306.2	51.0	138940	8	AC063973

ALIGNMENTS

RESULT 1
AF186184
LOCUS AF186184 8573 bp DNA linear PLN 10-SEP-2001
DEFINITION Glycine max retrovirus-like element Calypso3-1, partial sequence.
ACCESSION AF186184
VERSION AF186184.1 GI:6671124
KEYWORDS
SOURCE
ORGANISM
Glycine max.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.
REFERENCE
1 (bases 1 to 8573)
Wright,D.A. and Voytas,D.F.

Pred. No. is the number of results predicted by chance to have a

TITLE Calypso: A Heterogeneous Retrovirus-like Element Family from Glycine

JOURNAL max
REFERENCE Unpublished
2 (bases 1 to 8573)
AUTHORS Wright, D.A. and Voytas, D.F.
TITLE Direct Submission
JOURNAL Submitted (14-SEP-1999) Zoology and Genetics, Iowa State
University, 2208 Molecular Biology Bldg., Ames, IA 50011, USA

FEATURES
source
1. 8573
/organism="Glycine max"
/strain="L85"
/db_xref="taxon:3847"

repeat_region
1. 8452
/rpl_family="retrovirus-like element Calypso3-1"

gene
1. 3781
/gene="pol"

gene
4461. 5722
/gene="envelope-like"

misc_feature
7034. 7039
/note="polyurine tract"

LTR
BASE COUNT 2584 a 1731 c 2009 g 2249 t
ORIGIN

Query Match 95.28; Score 571.2; DB 8; Length 8573;
Best Local Similarity 97.0%; Pred. No. 8.6e-172;
Matches 582; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

1 TTGGAGGCTGGGCTCATATACCCCATCTGACAGCGCTTGGTAAGCCGATACAGG 60
1172 TTGGAGGCTGGGCTCATATACCCCATCTGACAGCGCTTGGTAAGCCGATACAGG 1231
61 GTTCCCAAGAAAGGTGAATGACAGTGTACAGATGAGAGAAATGACTTGATACCA 120
1232 GTTCCCAAGAAAGGTGAATGACAGTGTACAAATGAGAGAAATGACTTGATACCA 1291
121 CGAAGCTGCTAGTGGTGGCAATGTATGATGATGATGATGATGATGATGATGATG 180
1292 CGAAGCTGCTAGTGGTGGCAATGTATGATGATGATGATGATGATGATGATGATG 1351
181 AAGGACCATTTCCCTTACCTTTCATGATGATGATGATGATGATGATGATGATGATG 240
1352 AAGGACCATTTCCCTTACCTTTCATGATGATGATGATGATGATGATGATGATGATG 1411
241 TACTACTGTTCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
1412 TACTACTGTTCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1471
301 CAGAGAGAGAGCGCTTACATGCGCCCTTGGCGCTTGGCTTACAGAGAGATGATGAT 360
1472 CAGAGAGAGAGCGCTTACATGCGCCCTTGGCGCTTGGCTTACAGAGAGATGATGAT 1531
361 GGGTTATGATGACACGACCATTTTCAGAGATGATGATGATGATGATGATGATGATG 420
1532 GGGTTATGATGACACGACCATTTTCAGAGATGATGATGATGATGATGATGATGATG 1591
421 GTGGAGAGAGAGCATGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 480
1592 GTGGAGAGAGAGCATGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1651
481 AGCTGTTGAGAGCATGAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 540
1652 AGCTGTTGAGAGCATGAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1711
541 AATTGGAGAGAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
1712 AATTGGAGAGAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1771

RESULT 2
AF186182 10128 bp DNA linear PLN 05-JAN-2000

LOCUS AF186182
DEFINITION Glycine max retrovirus-like element Calypsol-1, partial sequence.
ACCESSION AF186182
VERSION AF186182.1 GI:66711122

KEYWORDS
SOURCE
ORGANISM
Glycine max.
Glycine max.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eustosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.

REFERENCE
1 (bases 1 to 10128)
Wright, D.A. and Voytas, D.F.
Calypso: A Heterogeneous Retrovirus-like Element Family from
Glycine max

AUTHORS
TITLE
JOURNAL
Unpublished
2 (bases 1 to 10128)
Wright, D.A. and Voytas, D.F.

REFERENCE
Wright, D.A. and Voytas, D.F.
Direct Submission
Submitted (14-SEP-1999) Zoology and Genetics, Iowa State
University, 2208 Molecular Biology Bldg., Ames, IA 50011, USA

FEATURES
source
1. 10128
Location/Qualifiers
/organism="Glycine max"
/strain="L85"
/db_xref="taxon:3847"

repeat_region
294. >10128
/rpl_family="retrovirus-like element Calypsol1-1"

LTR
primer_bind
294. 1598
1605. 1623
1759. 1712
/gene="pol"

gene
8068. 8976
/gene="envelope-like"

BASE COUNT 3031 a 2135 c 2330 g 2632 t
ORIGIN

Query Match 88.78; Score 532; DB 8; Length 10128;
Best Local Similarity 94.0%; Pred. No. 3.3e-159;
Matches 564; Conservative 0; Mismatches 35; Indels 1; Gaps 1;

1 TTGGAGGCTGGGCTCATATACCCCATCTGACAGCGCTTGGTAAGCCGATACAGG 60
4564 TTGGAGGCTGGGCTCATATACCCCATCTGATGATGATGATGATGATGATGATGATG 4623
61 GTTCCCAAGAAAGGTGAATGACAGTGTACAGATGAGAGAAATGACTTGATACCA 120
4624 GTTCCCAAGAAAGGTGAATGACAGTGTACAGATGAGAGAAATGACTTGATACCA 4683
121 CGAAGCTGCTAGTGGTGGCAATGTATGATGATGATGATGATGATGATGATGATGATG 180
4684 CGAAGCTGCTAGTGGTGGCAATGTATGATGATGATGATGATGATGATGATGATGATG 4743
181 AAGGACCATTTCCCTTACCTTTCATGATGATGATGATGATGATGATGATGATGATGATG 240
4744 AAGGACCATTTCCCTTACCTTTCATGATGATGATGATGATGATGATGATGATGATGATG 4803
241 TACTACTGTTCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
4804 TACTACTGTTCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4862
301 CAGAGAGAGAGCGCTTACATGCGCCCTTGGCGCTTGGCTTACAGAGAGATGATGAT 360
4863 CAGAGAGAGAGCGCTTACATGCGCCCTTGGCGCTTGGCTTACAGAGAGATGATGAT 4922
361 GGGTTATGATGACACGACCATTTTCAGAGATGATGATGATGATGATGATGATGATGATG 420
4923 GGGTTATGATGACACGACCATTTTCAGAGATGATGATGATGATGATGATGATGATGATG 4982

OY	421	GTGGAGAAACATCGAGGTAATTATTTAGAGACATCTTCOGTTTTTGACCCTCAATTGAC	480
Dd	4983	GTGGAGAGACATCGAGGTAATTATTTAGAGACACTTCTGGATTTTGGACCTCAATTGAC	504.2
OY	481	AGCTGTTTGGAGAACCTPAGAGAGGCTCTTCAGAGGTGCAGAGACTPAACTTGGTACTG	540
Dd	5043	AACATATTTGAGAACCTTAGAGATGCTACTACAGAGGTGCCATATGACTAATTGTTGTA	510.2
OY	541	AATGGGAAGAAAGTCAATTCATGCTTCAGAGGCGCATATGCTCAGGCCACAAGATCTCA	600
Dd	5103	AATTGGGAAGAAAGTCAATTCATGCTTCAGAGGCGCATATGCTCAGGCCACAAGATCTCA	516.2
RESULT 3			
AF378068			
LOCUS	AF378068	Glycine max isolate Soybean-3 retrovirus element Calypso, partial sequence.	
DEFINITION	AF378068	AF378068.1 GI:15724032	
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
FEATURES			
source			
BASE COUNT			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
OY	1	TTGAGAGCTGGGCTCATATATCCCATCTCTGACAGCGCTGGGTAGCCCATACAGSTG	60
Dd	25	TTTGAGAGTTGGGCTCATATATACCTCATCTCTGACAGCGCTGGGTAGCCCATACAGSTG	84
OY	61	GTTCCCAAGAAGAGTGGATGATGACAGTGTGACAGATGATGAGAGANTGACTGATACCA	120
Dd	85	GCTCCCAAGAAGATGCGGATATACAGTGTGACAAATGAGAGGAATGACTGATACCA	144
OY	121	CGAAGCTGCTACTGTTGGCGAATGTGATGACATATGCGACAGCTGAATGAAGCACACG	180
Dd	145	CGAAGCTGCACTGCTAGCGGATGTGATGACATGCTCAAGTTGAATGAAGCACACAG	204
OY	181	AAGAGCACTTCCCCCTTACCTTTATGATGATAGATGCTGAGAGACACTGACAGGCGAC	240
Dd	205	AAGAGCACTTCCCCCTTACCTTTATGATGATAGATGCTGAGAGAGCTTGGACGGCGAC	264

OY	241	TACACAGCTGTTCTGGATGATGACTCCGGGATACAAACCAAGATCGGGGTGACCCAGAAAT	300
Db	265	TACTACGTCTTCTGGATGATGATTCAGATATACAAACCAATCGGGGTGACCCAGAAAT	324
OY	301	CAGGAGAGAGAGCGCTTACATGCGCCCTTTGGCGCTTTTGCTTACAGAAAGATGCCATTTC	360
Db	325	CAGGAGAAAGATGCGCTTACATGCGCCCTTTGGCGCTTTTGCTTACAGAAAGATGCTATTTC	384
OY	361	GGGTATGTAATGCAACCAAGCCACATTTTCAGAGGTGATGCTGGCCATTTTTCAGACATG	420
Db	385	AGGTTATGTAATGCAACCAAGCCACATTTTCAGAGGTGATGCTGGCCATTTTTCAGACATG	444
OY	421	GTGGAGAAACATCGAGGTATTTATGAGACACTCTCGTGTGTTTGGACCCCTATTTGAC	480
Db	445	GTGGAGAAACATCGAGGTATTTATGAGACACTCTCGTGTGTTTGGACCCCTATTTGAC	504
OY	481	AGCTGTTTGAGAAACCTTAGAGAGGGTACTTCAGAGGTGCGAAGAGACTAACTTGATG	540
Db	505	AGTTGCTTAAAGCAACTTAGAGATGCTACTACAGAGGTGCGATAGACTAACTTGATG	564
OY	541	AATTGGGAAAGTGTCTATTCATGTTGTTGAGAGGGCATATGCTTAGGCCACACATCTCA	600
Db	565	AATTAGGAAATGTCTATTCATGTTGTTGAGAGGGCATATGAGTGGCCACAAATATCTCA	624
RESULT 4			
AF378062			
LOCUS	AF378062	762 bp	DNA linear PLN 11-JAN-2002
DEFINITION	Glycine max isolate Soybean8-2 retrovirus element Calypso, partial sequence.		
ACCESSION	AF378062		
VERSION	AF378062.1	GI:15724025	
KEYWORDS	Glycine max.		
SOURCE	Glycine max.		
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.		
REFERENCE	1 (bases 1 to 762)		
AUTHORS	Wright,D.A. and Voytas,D.F.		
TITLE	Athl14 of Arabidopsis and Calypso of soybean define a lineage of endogenous plant retroviruses		
JOURNAL	Genome Res. 12 (1), 122-131 (2002)		
MEDLINE	21638318		
PUBMED	11779837		
REFERENCE	2 (bases 1 to 762)		
AUTHORS	Voytas,D.F. and Wright,D.A.		
TITLE	Direct Submission		
JOURNAL	Submitted (08-MAY-2001) Department of Zoology and Genetics, Iowa State University, 2208 Molecular Biology, Ames, IA 50011, USA		
FEATURES	location/Qualifiers		
source	1..762		
	/organism="Glycine max"		
	/variety="L85"		
	/isolate="Soybean8-2"		
	/db_xref="taxon:3847"		
	/note="endogenous_virus: Calypso"		
	<1..>762		
misc_feature	/note="nonfunctional reverse transcriptase protein due to mutation"		
BASE COUNT	213 a	145 c	203 g
ORIGIN	201 t		
Query Match	87.7%: Score 526.4; DB 8; Length 762;		
Best Local Similarity	92.3%: Pred. No. 1.7e-157;		
Matches	554; Conservative	0; Mismatches	46; Indels
OY	1	TTGAGAGCTGGGCTCATATATACCCCATCTCTGACAGCGCTTGTTAGGCCAGTACAGTG	60
Db	25	TTGAGAGTTGGGCTCATATATACCTCTCTGACAGCGCTTGTTAGGCTAGTACAGTG	84
OY	61	GTTCOCCAAGAAAGTGCAATGCTGACAGATGAGAGAAATGACTTGATACCAACA	120

gene	7909..9171	/gene="envelope-like"
misc_feature	11223..11228	/pseudo
LTR	11232..12569	/note="polypurine tract"
BASE COUNT	3882 a 2456 c 2767 g 3781 t	
ORIGIN		
Query Match	87.2%;	Score 523.2; DB 8; Length 12886;
Best Local Similarity	93.2%;	Pred. Mismatches 38; Indels 3; Gaps 1;
Matches 559; Conservative 0;		
OY	1 TTGAGAGCGGGCTCATATATACCCCACTCTGACAGCGCTGGGTAAAGCCACTACAGTG 60	
DB	4293 TTGGAGGCTGGGCTCATATATACCCCTTCTCTACAGTGTGGTAAAGCCACTACAGTG 435	
OY	61 GTTCCCAAGAAAGTGGAATGACAGTGTGACAGATGAGAGAAATGACTTGTATACACA 120	
DB	4353 GTTCCCAAGAAAGTGGAATGACAGTGTGACAGAAATGAGAAATGACTTGTATACACA 441	
OY	121 CGAATGTCACGTGGTGGGAGATGTGTATCAGCTCTGCCAAGCTAAATGAAGCCACAGG 180	
DB	4413 CGAATTAATCACTGGTGGGAGATGTGTATCAATCAATGACCAAGCTAAATGAAGCCACAGA 447	
OY	181 AAGACCATTTCCCTTACCTTTCATGTGATCAGATGCTGAGAGACTTGACAGGCA 240	
DB	4473 AAGACCATTTCCCTTACCTTTCATGTGATCAGATGCTGAGAGACTTGAGGCA 453	
OY	241 TACTACGTCTTCTGGATGATACACGGGATACACAGATCGCGGTAGACCCAGAAAT 300	
DB	4533 TACTACATTTCTTGATGATACACGGGATATATACAGATCGCGGTAGACCCAGAAAT 459	
OY	301 CAGGAGAGACGGCCCTTACATGCCCCCTTGGCGCTTGTGCTTACAGAAAGATGCCATTTC 360	
DB	4593 CAGGAGAGCGGCGCTTACATGCCCCCTTGGCGCTTGTGCTTATAGAAAGATGCCATTTC 465	
OY	361 GGGTATATATATGACACCGACCAACATTTAGAGGTGATGCTGGCCATTTTTCACAGATG 420	
DB	4653 GGGTATATATATGACACCGACCAACATTTAGAGGTGATGCTGGCCATTTTTCACAGATG 471	
OY	421 GTGCGAAGAACATCGAGATATTTATATGACGACTCTCGGTTTTTGGACCCCATTTTGAC 480	
DB	4713 GTGAGAAAGAACATCGAGATATTTATATGACGACTCTCGGTTTTTGGACCCCATTTTAAAC 477	
OY	481 AGCTGTTGAGGAACCTTAGAGAGGGTACTTACAGAGGTGCGAAGACTTAACCTGTACTG 540	
DB	4773 A---GTTTGAAGAACCTTAGAGATGTTACTTATAGAGTTAGTAGAGACTTAACCTGTACTG 482	
OY	541 AATTGGGAAAAGTGCATTTTCATGTTGCGAAGGGCAATGTCCTTAGGCAACAATCTCA 600	
DB	4830 AACTGGGAAAAGTGCATTTTCATGTTGCGAAGGGCAATGTCCTTAGGCAACAATCTCA 488	
RESULT 6	763 bp DNA linear PLN 11-JAN-2002	
AF378063		
LOCUS	Glycine max isolate Soybean2 retrovirus element Calypso, partial	
DEFINITION	sequence.	
ACCESSION	AF378063	
VERSION	AF378063.1 GI:15724026	
KEYWORDS		
SOURCE	Glycine max.	
ORGANISM	Glycine max.	
REFERENCE	1 (bases 1 to 763)	
AUTHORS	Wright, D.A. and Voytas, D.F.	
TITLE	Achil4 of Arabidopsis and Calypso of soybean define a lineage of	
JOURNAL	endogenous Plant retroviruses	
	Genome Res. 12 (1), 122-131 (2002)	

MEDLINE 21638318
 PUBMED 11779837
 REFERENCE 2 (bases 1 to 763)
 AUTHORS Voytas,D.F. and Wright,D.A.
 TITLE Direct Submission
 JOURNAL Submitted (08-MAY-2001) Department of Zoology and Genetics, Iowa State University, 2208 Molecular Biology, Ames, IA 50011, USA

FEATURES
 source
 1..763
 /organism="Glycine max"
 /variety="L85"
 /isolate="Soybean2"
 /db_xref="taxon:3847"
 /note="endogenous_virus: Calypso"
 <1..>763
 /note="nonfunctional reverse transcriptase protein due to mutation"

misc_feature
 /note="nonfunctional reverse transcriptase protein due to mutation"

BASE COUNT 213 a 143 c 203 g 204 t

ORIGIN

Query Match
 Best Local Similarity 90.8%; Pred. No. 7.5e-150; Length 763;
 Matches 346; Conservative 0; Mismatches 54; Indels 1; Gaps 1;

1 TTGAGGCTGGGCTCATATACCCCATCTGACAGCGCTTGGGTAGCCGACGTACAGTG 60
 |||||
 25 TTGAGGCTGGGCTCATATATCCCATCTCTATATAGCACCTTAGAGCCGCTACAGTG 84
 |||||
 61 GTTCCCAAGAAAGGTGGAATGACAGTGTGTACAGAGTACAGAGAAATGACTGATACCA 120
 |||||
 85 GTTCCCAAGAAAGGTGGAATGACAGTGTGTACAGAGTACAGAGAAATGACTGATACCA 144
 |||||
 121 CGAAGCTGATGGTGGGGAATGTGATGATGATGATGATGATGATGATGATGATGATG 180
 |||||
 145 CGAAGCTGATGGTGGGGAATGTGATGATGATGATGATGATGATGATGATGATGATG 204
 |||||
 181 AAGGACCATTTCCCTTACCTTATGATGATGATGATGATGATGATGATGATGATGATG 240
 |||||
 205 AAGGACCATTTCCCTTACCTTATGATGATGATGATGATGATGATGATGATGATGATG 264
 |||||
 241 TACTACTGTTTCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 300
 |||||
 265 TATATATGTTTCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 324
 |||||
 301 CAGGAGAAAGCGGCTTACCTTATGATGATGATGATGATGATGATGATGATGATGATG 359
 |||||
 325 CAGGAGAAAGCGGCTTACCTTATGATGATGATGATGATGATGATGATGATGATGATG 384
 |||||
 360 CAGGAGAAAGCGGCTTACCTTATGATGATGATGATGATGATGATGATGATGATGATG 419
 |||||
 385 CAGGAGAAAGCGGCTTACCTTATGATGATGATGATGATGATGATGATGATGATGATG 444
 |||||
 420 GGTGGAAGAAAGCATGAGTATTTATGACGACTTCGGTTTTTGGACCCCTCATTTGA 479
 |||||
 445 GGTGGAAGAAAGCATGAGTATTTATGACGACTTCGGTTTTTGGACCCCTCATTTGA 504
 |||||
 480 CAGCTGTTTGGAGAACTTACCTTATGATGATGATGATGATGATGATGATGATGATG 539
 |||||
 505 CAGCTGTTTGGAGAACTTACCTTATGATGATGATGATGATGATGATGATGATGATG 564
 |||||
 540 GAATGGGAAAGGTGTCTATTTGATGATGATGATGATGATGATGATGATGATGATG 599
 |||||
 565 GAATGGGAAAGGTGTCTATTTGATGATGATGATGATGATGATGATGATGATGATG 624
 |||||
 600 A 600
 625 A 625

RESULT 7
 AF378064 762 bp DNA linear PLN 11-JAN-2002
 LOCUS AF378064
 DEFINITION Glycine max isolate soybean9-2 retrovirus element Calypso reverse

transcriptase gene, partial cds.
 AF378064 GI:15724027
 AF378064.1 GI:15724027

ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

REFERENCE
 AUTHORS Wright,D.A. and Voytas,D.F.
 TITLE Athlia4 of Arabidopsis and Calypso of soybean define a lineage of endogenous plant retroviruses
 JOURNAL Genome Res. 12 (1), 122-131 (2002)
 MEDLINE 21638318
 PUBMED 11779837
 REFERENCE 2 (bases 1 to 762)
 AUTHORS Voytas,D.F. and Wright,D.A.
 TITLE Direct Submission
 JOURNAL Submitted (08-MAY-2001) Department of Zoology and Genetics, Iowa State University, 2208 Molecular Biology, Ames, IA 50011, USA

FEATURES
 source
 1..762
 /organism="Glycine max"
 /variety="L85"
 /isolate="soybean9-2"
 /db_xref="taxon:3847"
 /note="endogenous_virus: Calypso"
 <1..>762
 /product="reverse transcriptase"
 /codon_start=1
 /product="reverse transcriptase"
 /protein_id="AL06417.1"
 /db_xref="GI:15724028"
 /translation="VRKEVILKLEVLGILYIPDSAWSPVLVYVSKREGMTVIRNEKND
 LIPRTVSMKICIDYRKLEINRKHDPFLPFDMDMERLAGAAVYCFIDAYGYNDI
 VYDRDQEKMATCPGCVFAYRIRIPGLCNAPPTTPMCKLAITADYKSEIEVPMDF
 SVFVPSLESLCKLEWLDRCVETNVLVNNKCHFPVRESIVLGHKISTGIEVDQK
 IDVIEKLPPSPSNVKGIRSLFGQAFYRFRFKIDPKY"

CDS
 mRNA
 BASE COUNT 228 a 150 c 181 g 203 t

ORIGIN

Query Match
 Best Local Similarity 82.4%; Pred. No. 1.5e-126; Length 762;
 Matches 493; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

2 TGGAGGCTGGGCTCATATACCCCATCTGACAGCGCTTGGGTAGCCGACGTACAGTG 61
 |||||
 26 TAGAGGTGGGCTTATATACCCCATCTCCGACAGCGCTTGGGTAGCCGACGTCTGGTG 85
 |||||
 62 TTCCCAAGAAAGGTGGAATGACAGTGTGTACAGAGTACAGAGAAATGACTGATACCA 121
 |||||
 86 TGTGGAAGAAAGGCGCATGACGATCATTCGAAATGAAAGAAATGACCTGATACCAAC 145
 |||||
 122 GAAGTGTCTACTGTTGGGCAATGTGTATGACATATGCAAGCTGAATGAAGCCACAG 181
 |||||
 146 GAAGTGTCTACTGTTGGGCAATGTGTATGACATATGCAAGCTGAATGAAGCCACAG 205
 |||||
 182 AGGACCATTTCCCTTACCTTATGATGATGATGATGATGATGATGATGATGATGATG 241
 |||||
 206 AGGACCATTTCCCTTACCTTATGATGATGATGATGATGATGATGATGATGATGATG 265
 |||||
 242 ACTACTGTTTCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 301
 |||||
 266 ATTATGCTTCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 325
 |||||
 302 AGGAGAAAGCGGCTTATATGATGATGATGATGATGATGATGATGATGATGATG 361
 |||||
 326 AGGAGAAAGCGGCTTATATGATGATGATGATGATGATGATGATGATGATGATG 385
 |||||
 362 GGTATGTATATGACACGACCATTTACAGAGTGTGATGATGATGATGATGATGATG 421
 |||||

```

Db      386 GGTGTGCAATGACACCAACATTCACAAATGTCATGTGGCCATTTTTCAGATATAG 445
OY      422 TGGAGAAAGACGAGGATTTATGAGACCTTCGGTTTGTGACCCCATTTGACA 481
Db      446 TGGAGAAAGACGAGGATTTATGAGACCTTCGATGATTTTGTGACCCCATTTGACA 505
OY      482 GCTGTTTGGAGAACCTAGAGAGGCTACTTCAGAGGTGGGAGAGACTAATGTTACTGA 541
Db      506 GTTGTTGAAGACGATGAGATGATACAAAGATGGGTGGAACAAATAGTACTA 565
OY      542 ATTGGAGAAAGCTGATTTATGTTTCAGAGAGGCTAGTCTAGGCCCAAGATGTC 599
Db      566 ATTGGAGAAAGCTGATTTATGTTTCAGAGAGGCTAGTCTAGGCCCAATTAATTC 623

RESULT 8
LOCUS   AP004896 92281 bp DNA linear PLN 19-JUL-2002
DEFINITION Lotus japonicus genomic DNA, chromosome 5, clone:LTJ16107, TM0040, complete sequence.
ACCESSION AP004896
VERSION   AP004896.1 GI:21907912
KEYWORDS  HG.
SOURCE    Lotus japonicus DNA, clone_L1b:LJT library clone:LTJ16107.
ORGANISM  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
          Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
          Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Lotaeae;
          Lotus.
REFERENCE 1
AUTHORS   Kameko,T., Nakamura,Y., Asamizu,E., Kato,T., Sato,S. and Tabata,S.
TITLE     Structural Analysis of a Lotus japonicus Genome. I. Sequence
          Features and Mapping of Sixty-six TAC clones which cover the 6.7 Mb
          Regions of the Genome
JOURNAL   Unpublished
REFERENCE 2 (bases 1 to 92281)
AUTHORS   Nakamura,Y.
TITLE     Direct Submission
          Submitted (26-MAR-2002) Yasukazu Nakamura, Kazusa DNA Research
          Institute, Department of Plant Gene Research, 1532-3, Yana,
          Kisarazu, Chiba 292-0812, Japan (E-mail:yn@kazusa.or.jp,
          URL:http://www.kazusa.or.jp, Tel:81-438-52-3935,
          Fax:81-438-52-3934)
FEATURES
   source          1..92281
                   /organism="Lotus japonicus"
                   /db_xref="taxon:34305"
                   /chromosome="5"
                   /clone_L1b="LJT library"
                   /note="TAC clone:TM0040"
BASE COUNT  29650 a 16291 c 17416 g 28924 t
ORIGIN
Query Match 59.8%; Score 359; DB 8; Length 92281;
Best Local Similarity 75.0%; Pred. No. 1.3e-103;
Matches 449; Conservative 0; Mismatches 150; Indels 0; Gaps 0;

OY      2 TGGAGGTGGGCTCATATACCCATCTGACAGCGCTGGTAAAGCCAGTACAGTGG 61
Db      78712 TAGAGCAGGATGATTTATCCATTTCTGATGATTTCTGGGTGACTCCAGTCAAGTGG 78771
OY      62 TTCCCAAGAAAGTGAATGACAGTGTACGAGATGAGAGGAATGACTTGATACCAAC 121
Db      78772 TTCCAAAGAAAGAGAGGATGATGATTTACTAATGACAAAGATGATGATTCGACAA 78831
OY      122 GAACGTCACTGCTGGTGGCAATGTGTATGATGATGCAAGCTGAATGAAGCCACACGA 181
Db      78832 GAACAGCATATGTTGGAGAAATGTGTATAGATTACAGGAAGCTGAACCAAGCCACGAA 78891
OY      182 AGGACATTTCCCTTACTTACATGATGATGATGCTGAGAGACTGAGAGGCGACGAT 241

```

```

Db      78892 AGGACATTTCCCATTTACATTCATGATGATGTTGGAGAGATTTCTGGGAGGCT 78951
OY      242 ACTACTGTTTCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 301
Db      78952 TCTATGCTTTCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 79011
OY      302 AGGAGAAAGCGGCTTTACATGATGATGATGATGATGATGATGATGATGATGATGATGATG 361
Db      79012 AGGAGAAAGCGGCTTTACATGATGATGATGATGATGATGATGATGATGATGATGATG 79071
OY      362 GGTATGTAATGACACCGCATTTTACAGAGGTGATGATGATGATGATGATGATGATGATGATG 421
Db      79072 GGTGTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 79131
OY      422 TGGAGAAAGACGAGGATTTATGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGA 481
Db      79132 TGGAGAAAGACGATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 79191
OY      482 GCTGTTTGGAGAACCTAGAGAGGCTACTTCAGAGGTGGGAGAGACTAATGTTGATGACA 541
Db      79192 TATGTTGCAAAAGCTTGACACCGCTGTGAAGCGGTGTGGAAACCAATCTGCTGCTTA 79251
OY      542 ATTGGAGAAAGCTGATTTATGTTTCAGAGAGGCTAGTCTAGGCCCAAGATGCTCA 600
Db      79252 ATTGGAGAAAGCTGATTTATGTTTCAGAGAGGCTAGTCTAGGCCCAATTAATTC 79310

RESULT 9
LOCUS   AF378072 762 bp DNA linear PLN 11-JAN-2002
DEFINITION Glycine max isolate soybean5-1 retrovirus element Calypso, partial
          sequence.
ACCESSION AF378072
VERSION   AF378072.1 GI:15724037
KEYWORDS  Glycine max.
SOURCE    Glycine max.
ORGANISM  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
          Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
          Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
          Glycine.
REFERENCE 1 (bases 1 to 762)
AUTHORS   Wright,D.A. and Voytas,D.F.
TITLE     Athila4 of Arabidopsis and Calypso of soybean define a lineage of
          endogenous plant retroviruses
JOURNAL   Genome Res. 12 (1), 122-131 (2002)
MEDLINE  21638318
PUBMED   11798337
REFERENCE 2 (bases 1 to 762)
AUTHORS   Voytas,D.F. and Wright,D.A.
TITLE     Direct Submission
          Submitted (08-MAY-2001) Department of Zoology and Genetics, Iowa
          State University, 2208 Molecular Biology, Ames, IA 50011, USA
FEATURES
   source          1..762
                   /organism="Glycine max"
                   /variety="Hark"
                   /isolate="Soybeans-1"
                   /db_xref="taxon:3847"
                   /note="endogenous_virus: Calypso"
BASE COUNT  224 a 132 c 186 g 220 t
ORIGIN
Query Match 58.8%; Score 352.6; DB 8; Length 762;
Best Local Similarity 74.3%; Pred. No. 1e-101;
Matches 445; Conservative 0; Mismatches 154; Indels 0; Gaps 0;

OY      2 TGGAGGTGGGCTCATATACCCATCTGTCAGAGCGCTGGTAAAGCCAGTACAGTGG 61
Db      26 TGGAGCAGGCTTATTTATTCATTTGATGATGATGATGATGATGATGATGATGATGATGATG 85

```



```

CDS
<1..>762
/product="reverse transcriptase"
/protein_id="FAL06418.1"
/db_xref="GI:15724035"
/translation="VRKEVEFLIEAGLILYIPISDSWVSPVOYVKKKGAMTVKNDRENE
LIPPRRYTRRMCIIDYRKINDEATLAKRKHDPILPEMOMKRLARLOSRYFLDQSGYNDI
AVDPDDEKTAFTCPPEFVYATRRRPFELCNASTTFQCKMAAIPDMVEKCLEYVNDPF
SFEFASGSPNCANLEKVLQRCENSLNLVNMWCKHFVMOEGIVLHSHKISKEIYEKKER
LDVILDKLPPPNVNVGHSIFLGHVGFYRFRFKDFKTV"
BASE COUNT      239 a      125 c      168 g      230 t
ORIGIN
Query Match      58.6%; Score 351.6; DB 8; Length 762;
Best Local Similarity 74.2%; Pred. No. 2.2e-101;
Matches 444; Conservative 0; Mismatches 154; Indels 0; Gaps 0;

OY      2 TGGAGGCGGGGCTCATATATACCCATCCTGACAGCGGTGGGTAAAGCCATACAGGCG 61
      1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      26 TAGAGGCGAGGGCTCATCTACCAATTTGAGTAGCTCTCGGTTATAGTCCGTTAAAGTTG 85
      1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY      62 TTCCCAAGAAAGGTGGAATGACAGTGTGATACAGATGAGAGGAATGACTTTATACAAAC 121
      1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      86 TTCCAAAAAAGAGAGGATGACAGTGTGATGATGATAGAAATAGATAGATATTTCTACAA 145
      1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY      122 GAACTGTACTGTGTGGCGAATGTGTATCGACTATCGCAAGCTGAATGAGCCACACGGA 181
      1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      146 GAAAGATCACAGATGAGAAATGTATGATTATAGAAACCTCAATGAAAGCCACAAAGAA 205
      1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY      182 AGGACCAATTCCTCCCTTACCTTCAATGGAATCGATGCGAGAGACTTGACAGGCGAGCAT 241
      1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      206 AAGACCAATTCCTCCCTTACCTTCAATGGAATCGATGCGAGAGACTTGCAAGGCGAATCT 265
      1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY      242 ACTACTGTTTCTTGATGATGATATCTCGGGATACAAACAGATGCGGATAGACCCAGAGATC 301
      1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      266 TCTACCGTTTCTTGAGCGGATACACAGGTTTACAAATGATATGAGATGCGATGCTCAGATC 325
      1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY      302 AGGAAAGACGGCCTTATCATGCCCCCTTGGCGCTTCTGTTACAGAAAGATGCCATTGC 361
      1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      326 AAGAAAAAAGCAGTTTACATGTCCTTTCAGCTGTTTGTGCTATGCGCGCATGGCGCTGC 385
      1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY      362 GGTATATGTAAGCAACGACCAATTTGAGAGGTGCATCGGCCATTTTTCAGACATGG 421
      1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      386 GTTATATGTAAGCCTTACTACTCTTACAGATGATATGATGAGCAATTTTATGATCATGG 445
      1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY      422 TGGAGAAAGCATGAGAGTATTTATGACGACTTCTCGTTTTTTTGGACCCATCTTGACA 481
      1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      446 TAGAGAAATGATTTGAAGTCTTTATGATGATGATTTTTCGTTCTTGTTGGTGCATCTTTGGAA 505
      1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY      482 GCTGTTGGAGAACCTAGAGAGGGTACTTCAAGAGGTGCGAAGACATGTTGGTACTGA 541
      1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      506 ATTGCTTAGCAAAATTTAGAGAAAGTGTTACCAACGTTTGAAAAAATCTAATTTGGGCTTGA 565
      1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY      542 ATTGGGAAAATGTCATCTTCAATGTTGTCAGAGGGCGATGTCCTAGGCCCAAGAATCTC 599
      1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      566 ACTGGGAAAAATGTCATCTTATGTTAGTACAAAGATGATGTCGTAGGACACAAAATCTC 623
      1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 11
AF378073      761 bp      DNA      linear      PLN 11-JAN-2002
LOCUS      AF378073
DEFINITION      Glycine max isolate Soybean5 retrovirus element Calypso, partial
sequence.
ACCESSION      AF378073
VERSION      AF378073.1
KEYWORDS      GI:15724038
SOURCE      Glycine max.
ORGANISM      Glycine max.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.
REFERENCE      1 (bases 1 to 761)

```

AUTHORS Wright, D.A. and Voytas, D.F.
 TITLE Athlia4 of Arabidopsis and Calypso of soybean define a lineage of
 endogenous plant retroviruses
 JOURNAL Genome Res. 12 (1), 122-131 (2002)
 MEDLINE 21638318
 PUBMED 11779837
 REFERENCE 2 (bases 1 to 761)
 AUTHORS Voytas, D.F. and Wright, D.A.
 TITLE Direct Submission
 JOURNAL Submitted (08-MAY-2001) Department of Zoology and Genetics, Iowa
 State University, 2208 Molecular Biology, Ames, IA 50011, USA

FEATURES
 source
 1..761
 /organism="Glycine max"
 /variety="Hark"
 /isolate="Soybeans"
 /db_xref="taxon:3847"
 /note="endogenous_virus: Calypso"
 misc_feature
 <1..>761
 /note="nonfunctional reverse transcriptase protein due to
 mutation"

BASE COUNT 217 a 130 c 185 g 229 t

ORIGIN

Query Match 58.3%; Score 349.8; DB 8; Length 761;
 Best Local Similarity 75.2%; Pred. No. 8.3e-101;
 Matches 449; Conservative 0; Mismatches 147; Indels 1; Gaps 1;

OY 2 TGGAGGCGGCGCTATATACCCCATCTGACAGCGCTGGGTAGCCGACATACAGTGG 61
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 26 TGGAGGAGGCGGCTATCTTATCTCATGATGATG-TTGGGTGATGTCAGTGCATGG 84
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 OY 62 TTCCCAAGAAAGGTGATGATGACAGTGTACGAGATGAGAGATGACTTGATACCAAC 121
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 85 TTCCCAAGAAAGGTGAGAGAGTGTGAGAGATGAGAGAAATAGACCTCATCTTACCC 144
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 OY 122 GAAGTGTACATGCTGTGGCGAATGTGTATGATGATGATGATGATGATGATGATGAT 181
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 145 GAAGTGTACATGCTGTGGCGAATGTGTATGATGATGATGATGATGATGATGATGAT 204
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 OY 182 AGGACCATTTCCCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTT 241
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 205 AGGACCATTTCCCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTT 264
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 OY 242 ACTACTGTTTCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 301
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 265 TCTATTTCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 324
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 OY 302 AGGAGAAAGCGGCTTACATGATGATGATGATGATGATGATGATGATGATGATGAT 361
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 325 AAGGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 384
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 OY 362 GGTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 421
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 385 AACTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 444
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 OY 422 TGGAGAAAGCGGCTTACATGATGATGATGATGATGATGATGATGATGATGATGAT 481
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 445 TGGAGAAAGCGGCTTACATGATGATGATGATGATGATGATGATGATGATGATGAT 504
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 OY 482 GCTGTTGAGGAGCTTACATGATGATGATGATGATGATGATGATGATGATGATGAT 541
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 505 GTTGTGAGGAGCTTACATGATGATGATGATGATGATGATGATGATGATGATGAT 564
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 OY 542 ATTGGAGAAAGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 598
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 565 ATTGGAGAAAGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 621
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 12
 AF378069
 LOCUS AF378069 761 bp DNA linear PLN 11-JAN-2002
 DEFINITION Glycine max isolate Soybean2 retrovirus element Calypso, partial

sequence.
 AF378069
 AF378069.1 GI:15724033
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Glycine max.
 Glycine max.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
 Glycine.

REFERENCE 1 (bases 1 to 761)
 AUTHORS Wright, D.A. and Voytas, D.F.
 TITLE Athlia4 of Arabidopsis and Calypso of soybean define a lineage of
 endogenous plant retroviruses
 JOURNAL Genome Res. 12 (1), 122-131 (2002)
 MEDLINE 21638318
 PUBMED 11779837
 REFERENCE 2 (bases 1 to 761)
 AUTHORS Voytas, D.F. and Wright, D.A.
 TITLE Direct Submission
 JOURNAL Submitted (08-MAY-2001) Department of Zoology and Genetics, Iowa
 State University, 2208 Molecular Biology, Ames, IA 50011, USA

FEATURES
 source
 1..761
 /organism="Glycine max"
 /variety="Williams"
 /isolate="Soybean2"
 /db_xref="taxon:3847"
 /note="endogenous_virus: Calypso"
 misc_feature
 <1..>761
 /note="nonfunctional reverse transcriptase protein due to
 mutation"

BASE COUNT 223 a 131 c 187 g 220 t

ORIGIN

Query Match 58.1%; Score 348.8; DB 8; Length 761;
 Best Local Similarity 74.3%; Pred. No. 1.7e-100;
 Matches 440; Conservative 0; Mismatches 152; Indels 0; Gaps 0;

OY 2 TGGAGGCGGCGCTATATACCCCATCTGACAGCGCTGGGTAGCCGACATACAGTGG 61
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 26 TGGAGGAGGCGGCTAT 85
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 OY 62 TTCCCAAGAAAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 121
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 86 TTCCCAAGAAAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 145
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 OY 122 GAAGTGTACATGCTGTGGCGAATGTGTATGATGATGATGATGATGATGATGATGAT 181
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 146 GAAGTGTACATGCTGTGGCGAATGTGTATGATGATGATGATGATGATGATGATGAT 205
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 OY 182 AGGACCATTTCCCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTT 241
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 206 AGGACCATTTCCCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTT 265
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 OY 242 ACTACTGTTTCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 301
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 266 ATTATGTTTCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 325
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 OY 302 AGGAGAAAGCGGCTTACATGATGATGATGATGATGATGATGATGATGATGATGAT 361
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 326 AAGGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 385
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 OY 362 GGTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 421
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 386 GTTGTGAGGAGCTTACATGATGATGATGATGATGATGATGATGATGATGATGAT 445
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 OY 422 TGGAGAAAGCGGCTTACATGATGATGATGATGATGATGATGATGATGATGATGAT 481
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 446 TGGAGAAAGCGGCTTACATGATGATGATGATGATGATGATGATGATGATGATGAT 505
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 OY 482 GCTGTTGAGGAGCTTACATGATGATGATGATGATGATGATGATGATGATGATGAT 541
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db	506	GGTGCTATTCACAACTTCTGAAGAGACTATTAAAGACACGTAAGAGATCCAAACTAGTTCTCA	565
Oy	542	ATTGGGAAAATGTCATTTTCATGTGTCGACAGGCATGCTCAGGCCACA	593
Db	566	ATTGGGAGAAATGTCATTTTCATGTGTCGACAGGAAATGCTGGGCGCATATAA	617
RESULT 13			
ACI23573/c			
LOCUS	ACI23573	110729 bp	DNA linear HTG 03-AUG-2002
DEFINITION	Medicago truncatula clone mth2-5n3, WORKING DRAFT SEQUENCE, 12		
ACCESSION	ACI23573		
VERSION	ACI23573.8	GI:22094376	
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.		
SOURCE	dartel medic.		
ORGANISM	Medicago truncatula		
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
	Spermatophytia; Magnoliophyta; eudicotyledons; core eudicots;		
	Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;		
	Medicago.		
REFERENCE	1 (bases 1 to 110729)		
AUTHORS	Shauli,S., Ilan,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,		
TITLE	Cook,D., Kim,D. and Roe,B.A.		
JOURNAL	Medicago truncatula BAC Clone mth2-5n3		
REFERENCE	unpublished		
AUTHORS	2 (bases 1 to 110729)		
TITLE	Shauli,S., Ilan,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,		
JOURNAL	Cook,D., Kim,D. and Roe,B.A.		
	Submitted (31-MAY-2002) Department Of Chemistry And Biochemistry,		
	The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,		
	OK 73019, USA		
	Submitted (03-AUG-2002) Department Of Chemistry And Biochemistry,		
	The University Of Oklahoma, 620 Farrington Oval, Room 208, Norman,		
	OK 73019, USA		
COMMENT	On Aug 3, 2002 this sequence version replaced gi:22002169.		
	----- Genome Center -----		
	Center: Department Of Chemistry And Biochemistry		
	The University Of Oklahoma		
	Center code:OOkNOR		

*	* NOTE: This is a 'working draft' sequence. It currently		
*	* consists of 12 contigs. The true order of the pieces		
*	* is not known and their order in this sequence record is		
*	* arbitrary. Gaps between the contigs are represented as		
*	* runs of N, but the exact sizes of the gaps are unknown.		
*	* This record will be updated with the finished sequence		
*	* as soon as it is available and the accession number will		
*	* be preserved.		
*	1		
*	3190: contig of 3190 bp in length		
*	3191	3290: gap of unknown length	
*	3291	6614: contig of 3324 bp in length	
*	6615	6714: gap of unknown length	
*	6715	14583: contig of 7869 bp in length	
*	14584	14683: gap of unknown length	
*	14684	23214: contig of 8531 bp in length	
*	23215	23414: gap of unknown length	
*	23415	28473: contig of 5159 bp in length	
*	28474	28573: gap of unknown length	
*	28574	38961: contig of 10388 bp in length	
*	38962	39061: gap of unknown length	
*	39062	46313: contig of 7252 bp in length	
*	46314	46413: gap of unknown length	
*	46414	53821: contig of 7408 bp in length	
*	53822	53921: gap of unknown length	
*	53922	65950: contig of 12029 bp in length	
*	65951	66050: gap of unknown length	
*	66051	80887: contig of 14837 bp in length	

FEATURES	source	1. 110729	/organism="Medicago truncatula"	/db_xref="taxon:3880"	/clone="mch2-5n3"	/clone_lib="Medicago truncatula BAC library H2"
* 80888	80987: gap of unknown length					
* 80988	92571: contig of 11554 bp in length					
* 92572	92671: gap of unknown length					
* 92672	110729: contig of 18058 bp in length.					
	Location/Qualifiers					
	1. 110729					
ORIGIN						
BASE COUNT	34866 a	18674 c	20249 g	35833 t	1107 others	
Query Match	57.7%;	Score 346.2;	DB 2;	Length 110729;		
Best Local Similarity	73.6%;	Pred. No. 1.7e-99;				
Matches 441; Conservative	0;	Mismatches 158;	Indels 0;	Gaps 0;		
QY	2	TGGAGCGTGGGTCATATATACCCCATCTCAGCGCGCTGGGTAACCCAGTACAGTGG	61			
Db	63722	TTGAGACCGGTATGATGATTCATCCATCTCGGATGCAAAATGGTAAATCCGGTGCACTGG	63666			
QY	62	TTCCCAAGAAAGGTGGAATGACAGTGGTACAGATGAGAGATGACTTGAATACCAACAC	121			
Db	63662	TGCGCAAGAAAGGGTGGCATGACCGGTATGTCAAAATAGAAAGATGAGTAAATTCCTAC	63603			
QY	122	GAACTGTCACTGGTTGGCCATGTGATTCGACTATGCAAGCTGAATGAAGCCACACGA	181			
Db	63602	GAACCGTCACCGGGTGGCGGATGTGATGATTCAGAAAGACTTAAACCAAGCTACAAAG	63543			
QY	182	AGGACACTTTCGCCCTTACCTTTGATGGATGAGATGCTGGAGAGACTTGAAGGGCAGCAT	241			
Db	63542	AAGATCTATTTCCCATCTGCTTATGATGACCAAAATGCTTGAGCGCTTTCGGTCAAGCTT	63488			
QY	242	ACTACTGTCTTCTTGATGATGATCTCGGGATACCAACGATCGCGGTAGACCCCAAGATC	301			
Db	63482	ATTATTTCTTCTTGATGATGCTATTCGGGCTACCAACCAATACAGTCAATCGGTTGATG	63422			
QY	302	AGGAGAAGAGCGCCTTTACATGCCCCCTTTGGCGTCTTTGTTACAAAGATGCCATTCG	361			
Db	63422	ATGGAAGAGCGGCTTTTACATGTCCTTTTGGAATCTTTGCTTATTAAGAAATGCCGTTG	63363			
QY	362	GGTATGATTAAGACACGACCAATTTACAGAGTGATGATGCGCATTTTTCAGACATG	421			
Db	63362	GTTTATTCATTCGCGCTCGCAACATTTCCAAAGATGATGCAAGGATATTTCTCGATTTGA	63303			
QY	422	TGGAAGAAAGCATCGAGGTATTTATGAGACGACTTCTCGTTTTTGGACCCCTATTGGACA	481			
Db	63302	TTGGAAGAGATATTTAGAGGTATTCATGAGTATTTCTCTGTTTTTGGATCATCCTATGAG	63243			
QY	482	GCTGTTTGAAGAACCTTAGAGAGGGTACTTCAGAGGTGCGAGAGACTTAATCTTGTACTGA	541			
Db	63242	TTTGCTTTGAACAACCTTGAGACCGTGTCTGAAAGCGGGTAAAGAAACCAATCTTGTCCTTA	63183			
QY	542	ATTGGGAAAGTGATTCATTCATGCTTGCAGAGAGGGCATATGCCATAGGCCCAAGATGTCA	600			
Db	63182	ATTGGGAAAGTGCATTTTCATGCTGTACCGAGGGCATTTGTTCTTGCTACAAAGATTTCA	63124			
RESULT 14						
AF186185						
LOCUS	AF186185	13637 bp	DNA	linear	PLN 10-SEP-2001	
DEFINITION	Glycine max retrovirus-like element Calypso4-1, partial sequence.					
ACCESSION	AF186185					
VERSION	AF186185.1	GI:6671125				
KEYWORDS						
SOURCE						
ORGANISM	Glycine max. Glycine max. Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eustosids I.; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.					
REFERENCE	1 (bases 1 to 13637) Wright,D.A. and Voytas,D.F.					
AUTHORS						

TITLE Calypso: A Heterogeneous Retrovirus-like Element Family from Glycine max

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 13637)

AUTHORS Wright, D.A. and Voytas, D.F.

TITLE Direct Submission

JOURNAL Submitted (14-SEP-1999) Zoology and Genetics, Iowa State University, 2208 Molecular Biology Bldg., Ames, IA 50011, USA

FEATURES

SOURCE 1. 13637

ORGANISM="Glycine max"

/strain="L85"

/db_xref="taxon:3847"

repeat_region 3335..12472

/rpt_family="retrovirus-like element Calypso4-1"

gene 3335..7628

/gene="pol"

gene 8205..9350

/pseudo

/gene="envelope-like"

/pseudo 11100..11106

misc_feature 11109..12472

/note="polyurine tract"

LTR 11109..12472

BASE COUNT 4227 a 2455 c 2912 g 4043 t

ORIGIN

Query Match 57.6%; Score 345.6; DB 8; Length 13637;

Best Local Similarity 73.5%; Pred. No. 2.3e-99;

Matches 441; Conservative 0; Mismatches 159; Indels 0; Gaps 0;

1 TTGGAGGCTGGGCTCATATACCCCATCTCTGACAGCGCTGGTAAGCCAGTACAGTG 60

4979 TAGAAGAGAGGCGCTATTATACCCCTCCGATGATGCGATGGGTTAGCCTTGCGAGGTT 5038

61 GTTCCCAAGAAAGGTGATGACAGTGGTACGATGAGAGAGATGACTTGATCCACAA 120

5039 GTCCCAAGAAAGGTGATGACAGTGGTACGATGAGAGAGATGACTTGATCCACAA 5098

121 CGAAGTGTACCTGGTGGGAGATGATGATGACATGCGACGATGAGAGAGATGAGAGAG 180

5099 AGGACTGTACCGGGGTGAGATGATGATGACATGCGACGATGAGAGAGATGAGAGAG 5158

181 AAGGACATTTCCCTTACCTTTCATGATGATGATGATGATGATGATGATGATGATGAT 240

5159 AAGGACATTTCCCTTACCTTTCATGATGATGATGATGATGATGATGATGATGATGAT 5218

241 TACTACTGTTCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300

5219 TATTATTGTTTCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5278

301 CAGGAGAAAGCGGCTTACATGCGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGG 360

5279 CAGGAGAAAGCGGCTTACATGCGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGG 5338

361 GGGTTATGATGACACGACATTTGAGAGGTCATGCTGCGCATTTTTCAGACATG 420

5339 GGGTTATGATGACACGACATTTGAGAGGTCATGCTGCGCATTTTTCAGACATG 5398

421 GTGAGAAAGACATGAGAGTATTATGACGACCTTCTCGGTTTGGACCCCTCATTTGAC 480

5399 GTGAGAAAGACATGAGAGTATTATGACGACCTTCTCGGTTTGGACCCCTCATTTGAC 5458

481 AGCTGTTGAGAGACCTGAGAGGCTGATGAGAGTGGAGAGAGCTAAGTGGTACG 540

5459 GGGTGGCTATTAACTTGAAGAGTATTACAGAGATGAGAGAGTCCATCTAGTTTC 5518

541 AATGAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600

5519 AATGAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5578

RESULT 15

AF378071 761 bp DNA linear PLN 11-JAN-2002

LOCUS Glycine max isolate Soybean2 retrovirus element Calypso, partial sequence.

DEFINITION

ACCESSION AF378071

VERSION AF378071.1

KEYWORDS GI:15724036

SOURCE

ORGANISM

Glycine max.

Eukaryote: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseolae;

REFERENCE 1 (bases 1 to 761)

AUTHORS Wright, D.A. and Voytas, D.F.

TITLE Athila4 of Arabidopsis and Calypso of soybean define a lineage of endogenous plant retroviruses

JOURNAL Genome Res. 12 (1), 122-131 (2002)

MEDLINE 21638318

PUBMED 1179837

REFERENCE 2 (bases 1 to 761)

AUTHORS Voytas, D.F. and Wright, D.A.

TITLE Direct Submission

JOURNAL Submitted (08-MAY-2001) Department of Zoology and Genetics, Iowa State University, 2208 Molecular Biology, Ames, IA 50011, USA

FEATURES

SOURCE 1. 761

ORGANISM="Glycine max"

/variety="Hark"

/isolate="Soybean2"

/db_xref="taxon:3847"

/note="endogenous_virus: Calypso"

/note="nonfunctional reverse transcriptase protein due to mutation"

BASE COUNT 224 a 134 c 180 g 223 t

ORIGIN

Query Match 57.0%; Score 342.2; DB 8; Length 761;

Best Local Similarity 74.3%; Pred. No. 2.3e-98;

Matches 445; Conservative 0; Mismatches 153; Indels 1; Gaps 1;

2 TGGAGGCTGGGCTCATATACCCCATCTCTGACAGCGCTGGTAAGCCAGTACAGTG 61

26 TAGAAGAGAGGCGCTATTATACCCCTCCGATGATGCGATGGGTTAGCCTTGCGAGGTTG 85

62 TTCCCAAGAAAGGTGATGACAGTGGTACGATGAGAGAGATGACTTGATCCACAA 121

86 TCCCGAAGAAAGGTGATGACAGTGGTACGATGAGAGAGATGACTTGATCCACAA 145

122 GAACTGTACCTGGTGGGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 181

146 GAACTGTACCGGGT-GAAGATGTCATGATGATGATGATGATGATGATGATGATGATGAT 204

182 AGGACATTTCCCTTACCTTTCATGATGATGATGATGATGATGATGATGATGATGATGAT 241

205 AGGACATTTACCTTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 264

242 ACTACTGTTTCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 301

265 ATTATGTTTCTTAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 324

302 AGGAGAGAGCGGCTTACATGCGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGG 361

325 AGGAGAGAGCGGCTTACATGCGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGG 384

362 GGTATGATGATGACACGACATTTGAGAGGTCATGCTGCGCATTTTTCAGACATG 421

385 GCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 444

422 TGGAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 481

Sat Jun 21 17:38:36 2003

us-09-965-553-11.rge

Page 11

Db 445 TGGAAAAATGCATTGACCTTTTCATGGACGANTTTCATTTTGGGCACATCTTTGGAG 504
 Oy 482 GCTGTTTAGGAGCACTAGACGGGTACTTCACAGGCGCAAGACATACTGCTACTGA 541
 Db 505 GCTACTTTCACAACTGTGAAGAGATTTACAGAGATGTACAGCTTAATCTTGTTCTCA 564
 Oy 542 ATTGGGAAGATGTCATTTTCATGCTTCGAGAGAGCGCATATCTTCAGGCCACAGATCTCA 600
 Db 565 ATTGGAGAAATGTCATTTCATGCTTCAGAGAGATGATGCTGGGGGCATATAAATTTCA 623

```
Search completed: June 20, 2003, 23:24:00
Job time : 2522.09 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2003, 18:46:16 ; Search time 343.396 Seconds
(without alignments)
3934.810 Million cell updates/sec

Title: US-09-965-553-11

Perfect score: 600

Sequence: 1 ttgagagctgtgcctacata.....tccatgacacaaatctca 600

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

N.Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	600	100.0	600	21	AAZ35258
2	600	100.0	12286	21	AAZ35261
3	571.2	95.2	4609	21	AAZ35273
4	568.2	94.7	597	21	AAZ35280
5	532	88.7	9829	21	AAZ35271
6	523.2	87.2	12571	21	AAZ35272
7	345.6	57.6	9139	21	AAZ35274
8	313.8	52.3	597	21	AAZ35266
9	307.8	51.3	94895	21	AAZ35262

10	307.8	51.3	1082138	21	AAZ22305	Arabidopsis thaliana
11	304.6	50.8	600	21	AAZ35264	Arabidopsis thaliana
12	304.6	50.8	59590	21	AAZ22281	BAC containing rep
13	303	50.5	83590	21	AAZ22283	BAC containing rep
14	303	50.5	90336	21	AAZ22289	BAC containing rep
15	303	50.5	94895	21	AAZ22302	BAC containing rep
16	303	50.5	103929	21	AAZ22287	BAC containing rep
17	303	50.5	134499	21	AAZ22286	BAC containing rep
18	303	50.5	1082138	21	AAZ22305	Arabidopsis thaliana
19	301.4	50.2	79122	21	AAZ22294	BAC containing rep
20	300	50.0	92584	21	AAZ22288	BAC containing rep
21	295.8	49.3	129021	21	AAZ22296	BAC containing rep
22	295	49.2	163119	21	AAZ22306	Arabidopsis thaliana
23	291.4	48.6	64415	21	AAZ22279	BAC containing rep
24	287.8	48.0	64415	21	AAZ22279	BAC containing rep
25	287	47.8	109973	21	AAZ22298	BAC containing rep
26	274.2	45.7	611590	21	AAZ22203	Arabidopsis thaliana
27	273.4	45.6	40349	21	AAZ22278	BAC containing rep
28	272	45.3	96988	21	AAZ22290	BAC containing rep
29	251.6	41.9	80450	21	AAZ22295	BAC containing rep
30	172.4	28.7	10482	21	AAZ35275	Soybean retroelement
31	159.2	26.5	109973	21	AAZ22298	BAC containing rep
32	102.8	17.1	91552	24	AAZ38803	BAC clone K6P36 fr
33	97.2	16.2	94618	21	AAZ22285	BAC containing rep
34	97	16.2	8435	20	AAZ3445	Genomic sequence o
35	91.4	15.2	82588	21	AAZ22301	BAC containing rep
36	89.4	14.9	90336	21	AAZ22289	BAC containing rep
37	89.4	14.9	121001	21	AAZ22284	BAC containing rep
38	86	14.3	101786	21	AAZ22293	BAC containing rep
39	85.6	14.3	59590	21	AAZ22281	BAC containing rep
40	85.6	14.3	151826	21	AAZ22291	BAC containing rep
41	84.2	14.0	4271	23	ABL04659	Drosophila melanog
42	84.2	14.0	6535	23	ABL04658	Drosophila melanog
43	82.2	13.7	163119	23	ABL22306	Arabidopsis thaliana
44	81	13.5	20213	23	ABL22121	Drosophila melanog
45	74.8	12.5	96583	21	AAZ22297	BAC containing rep

ALIGNMENTS

RESULT 1
AAZ35258
ID AAZ35258 standard; DNA; 600 BP.

AC AAZ35258;
XX
DT 27-MAR-2000 (first entry)
XX
DE Plant retroelement generic reverse transcriptase gene.
XX
KW Retroelement; retrovirus; transgenic plant; gene transfer;
KW soybean; pea; reverse transcriptase; Calypso; athalia; cyclops; ss.
XX
OS Glycine max.
OS Arabidopsis thaliana.
OS Pisum sativum.
XX
PN WO960842-A2.
XX
PD 02-DEC-1999.
XX
PF 28-MAY-1999; 99WO-US11858.
XX
PR 29-MAY-1998; 98US-0087125.
PR 28-MAY-1999; 99US-0087125.
PA (WRIG/) WRIGT D A.
PA (VOYT/) VOYTAS D F.
XX
PI Wright DA, Voytas DF;
XX
DR WPI; 2000-105586/09.

DR P-PSDB; AAU32431.

XX New nucleic acid molecules for imparting agronomically significant
PT characters to plants, especially soybean

XX Claim 1(e); Page 80; 118pp; English.

CC The present sequence comprises a consensus of reverse transcriptase
CC gene sequences identified in retrovirus-like element (retroelement)
CC calypso of soybean (see AA235280), cyclops of pea (see AA235266) and
CC athlia of Arabidopsis thaliana (see AA235264). The invention
CC provides molecular tools in the form of retroelements and
CC retroelement-containing vectors, cells and plants. Methods are
CC provided for introducing the retroelements into cells, especially
CC when the retroelement carries at least 1 agronomically-significant
CC characteristic (ACS). In a preferred method, a helper cell line
CC which expresses gag, pol and env sequences is used to enable
CC transfer of a secondary construct which carries an ASC and has
CC retroelement sequences that allow for replication and integration.
CC Claimed isolated nucleic acid molecules comprise a nucleic acid
CC sequence selected from a retroelement primer binding site,
CC envelope, gag, integrase, reverse transcriptase, protease or
CC RNAse-H sequence (see AA235254-61). Also provided are plant
CC retroviral particles that are used to transfer the nucleic acids
CC into plant cells.

XX Sequence 600 BP; 157 A; 128 C; 166 G; 149 T; 0 other;

Query Match 100.0%; Score 600; DB 21; Length 600;

Best Local Similarity 100.0%; Pred. No. 6.5e-193;

Matches 600; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGAGGCTGGGCTCATATACCCATCTCTGACAGCGCTTGGTAAAGCCAGTACAGTG 60
DB 1 TTGAGGCTGGGCTCATATACCCATCTCTGACAGCGCTTGGTAAAGCCAGTACAGTG 60
QY 61 GTTCCCAAGAAAGGTGGAATGACAGTGTGACAGATGAGAGGAATGACTTGTATCCACA 120
DB 61 GTTCCCAAGAAAGGTGGAATGACAGTGTGACAGATGAGAGGAATGACTTGTATCCACA 120
QY 121 CGAAGTGTACTGTGGCGAATGTATGACACTTCGCAAGCTGAATTAACCCACAGG 180
DB 121 CGAAGTGTACTGTGGCGAATGTATGACACTTCGCAAGCTGAATTAACCCACAGG 180
QY 181 AAGGACCATTTCCCTTACCTTTATGATAGATGCTGAGAGAGACTTGCAGGAGGCA 240
DB 181 AAGGACCATTTCCCTTACCTTTATGATAGATGCTGAGAGAGACTTGCAGGAGGCA 240
QY 241 TACTACTGTTTCTTGATGATGATCGGATACCAACGATCGCGTGAACCCAGAGAT 300
DB 241 TACTACTGTTTCTTGATGATGATCGGATACCAACGATCGCGTGAACCCAGAGAT 300
QY 301 CAGGAGAAAGCGGCTTACATGCCCCCTTGGCGTTTGTACAGAAAGATGCAATTC 360
DB 301 CAGGAGAAAGCGGCTTACATGCCCCCTTGGCGTTTGTACAGAAAGATGCAATTC 360
QY 361 GGGTATATGATGACCAAGCAATTCAGAGGTGATGCTGGCCATTTTTCAGACATG 420
DB 361 GGGTATATGATGACCAAGCAATTCAGAGGTGATGCTGGCCATTTTTCAGACATG 420
QY 421 GTTGAGAGAAAGCATGAGATATTATGAGACATCTTCGCTTTTGGACCCCTCATTTTAC 480
DB 421 GTTGAGAGAAAGCATGAGATATTATGAGACATCTTCGCTTTTGGACCCCTCATTTTAC 480
QY 481 AGCTGTTTGAAGAACTAGAGAGGCTACTTCAGAGTGCAGAAAGACTAATCTGTACTG 540
DB 481 AGCTGTTTGAAGAACTAGAGAGGCTACTTCAGAGTGCAGAAAGACTAATCTGTACTG 540
QY 541 AATTGGGAAAAGTGTCTATTCATGATGTTGAGAGGGCATATGCTTGAAGCCACAGATCTCA 600
DB 541 AATTGGGAAAAGTGTCTATTCATGATGTTGAGAGGGCATATGCTTGAAGCCACAGATCTCA 600

RESULT 2

AA235261
ID AA235261 standard; DNA; 12286 BP.

XX AA235261;

XX 27-MAR-2000 (first entry)

XX Plant generic retroelement.

XX Retroelement; retrovirus; transgenic plant; gene transfer;

XX soybean; pea; Calypso; athlia; cyclops; ss.

XX Glycine max.

XX Arabidopsis thaliana.

XX Pisum sativum.

XX Key Location/Qualifiers

XX CDS 1482..6887

XX W09960842-A2.

XX 28-MAY-1999; 99W0-US11858.

XX 29-MAY-1998; 98US-0087125.

XX 28-MAY-1999; 99US-0087125.

XX (WRIG/) WRIGHT D A.

XX (VOYT/) VOYTAS D F.

XX Wright DA, Voytas DF;

XX WPI; 2000-105586/09.

XX P-PSDB; AAU32434.

XX New nucleic acid molecules for imparting agronomically significant

XX characters to plants, especially soybean

XX Claim 1(h); Page 84-88; 118pp; English.

CC The present sequence comprises a generic plant retroelement
CC obtained from retrovirus-like elements (retroelements) calypso of
CC soybean, cyclops of pea and athlia of Arabidopsis thaliana. The
CC invention provides molecular tools in the form of retroelements and
CC retroelement-containing vectors, cells and plants. Methods are
CC provided for introducing the retroelements into cells, especially
CC when the retroelement carries at least 1 agronomically-significant
CC characteristic (ACS). In a preferred method, a helper cell line
CC which expresses gag, pol and env sequences is used to enable
CC transfer of a secondary construct which carries an ASC and has
CC retroelement sequences that allow for replication and integration.
CC Claimed isolated nucleic acid molecules comprise a nucleic acid
CC sequence selected from a retroelement primer binding site,
CC envelope, gag, integrase, reverse transcriptase, protease or
CC RNAse-H sequence (see AA235254-61). Also provided are plant
CC retroviral particles that are used to transfer the nucleic acids
CC into plant cells.

XX Sequence 12286 BP; 3748 A; 2540 C; 2767 G; 3331 T; 0 other;

Query Match 100.0%; Score 600; DB 21; Length 12286;

Best Local Similarity 100.0%; Pred. No. 3.5e-192;

Matches 600; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGAGGCTGGGCTCATATACCCATCTCTGACAGCGCTTGGTAAAGCCAGTACAGTG 60
DB 4278 TTGAGGCTGGGCTCATATACCCATCTCTGACAGCGCTTGGTAAAGCCAGTACAGTG 4337
QY 61 GTTCCCAAGAAAGGTGGAATGACAGTGTGACAGATGAGAGGAATGACTTGTATCCACA 120
DB 61 GTTCCCAAGAAAGGTGGAATGACAGTGTGACAGATGAGAGGAATGACTTGTATCCACA 120

Db	4338	GTTCCCAAGAAAGGTGGAAATGACAGTGTGACAGATGAGAGAAATGACATTGATACCAACA	43397
QY	121	CGAACGTCTACTGTTGGCGANGTGTATGACTATCGCAAGCTGAATGAAGCCACACGG	180
Db	4398	CGAACTGTCACTGGTTGGCGAAGTGTATGACTATCGCAAGCTGAATGAAGCCACACGG	4457
QY	181	AAGACACATTTCCTTACCTTTTCATGATGCATGCTGGAGATCACTTGCAGGGCAGGCA	240
Db	4458	AAGACACATTTCCTTACCTTTTCATGATGCATGCTGGAGATCACTTGCAGGGCAGGCA	4517
QY	241	TACTACTGTTTCTTGATGGATGCTCGGGATACCAACAGATCGGGTAGACCCAGAGAT	300
Db	4518	TACTACTGTTTCTTGATGGATGCTCGGGATACCAACAGATCGGGTAGACCCAGAGAT	4577
QY	301	CAGGAGAAAGACGGGCTTTTAAATGGCCCTTTGGGCGCTTTGGTTACAGAAAGATCCATTC	360
Db	4578	CAGGAGAAAGACGGGCTTTTAAATGGCCCTTTGGGCGCTTTGGTTACAGAAAGATCCATTC	4637
QY	361	GGGTTATGTATGATGACACGACGACATTTTCAGAGTGCATGCTGGCCATTTTTCAGACATG	420
Db	4638	GGGTTATGTATGATGACACGACGACATTTTCAGAGTGCATGCTGGCCATTTTTCAGACATG	4697
QY	421	GTGGAGAAAGCAATCGAGATTTATATGAGACACTTCGCTTTTGGACCTCATTTGAC	480
Db	4698	GTGGAGAAAGCAATCGAGATTTATATGAGACACTTCGCTTTTGGACCTCATTTGAC	4757
QY	481	AGCTGTTTGGAGAACTAGAGAGGGTACTTCAGAGGTGCGAAGAGACTAACTTGGTACTG	540
Db	4758	AGCTGTTTGGAGAACTAGAGAGGGTACTTCAGAGGTGCGAAGAGACTAACTTGGTACTG	4817
QY	541	AATTGGGAAAAGTCTCATTTTCATGCTTCGAGAGGGCATAGTCTCTAAGCCACAAATCTCA	600
Db	4818	AATTGGGAAAAGTCTCATTTTCATGCTTCGAGAGGGCATAGTCTCTAAGCCACAAATCTCA	4877

XX	AA235273	
XX	AA235273	standard; DNA; 4609 BP.
XX	AA235273;	
XX	27-MAR-2000	(first entry)
XX	Soybean retroelement Calypso 1-3.	
XX	DE	
XX	Retroelement; retrovirus; transgenic plant; gene transfer;	
XX	KW	
XX	KM	Calypso 1-3; soybean; ss.
XX	XX	
XX	OS	Glycine max.
XX	PN	
XX	W09960842-A2.	
XX	PD	
XX	02-DEC-1999.	
XX	28-MAY-1999;	99MO-US11858.
XX	PE	
XX	29-MAY-1998;	98US-0087125.
XX	PR	
XX	28-MAY-1999;	99US-0087125.
XX	PA	
XX	(WRIGHT) WRIGHT D A.	
XX	PA	
XX	(VOYT/) VOYTAS D F.	
XX	PI	
XX	Wright DA,	Voytas DF;
XX	DR	
XX	WPI: 2000-105586/09.	
XX	PT	
XX	PS	
XX	Example 3;	Page 102-104; 118pp; English.

This is the nucleotide sequence of the Calypso 1-3 retroelement of soybean. It was identified by screening of a soybean lambda ccosyban. It was identified by screening of a soybean lambda

CC library using a reverse transcriptase probe. 2 Groups of soybean
CC retroelements were identified, i.e. Calypso 1-1, 1-2 and 1-3
CC (see AA352271-73) and Calypso 2-1 and 2-2 (see AA35274-75). The
CC retroelements include gag, pol, env and primer binding site
CC sequences that can be used in constructs of the invention. The
CC invention provides molecular tools in the form of retroelements and
CC retroelement-containing vectors, cells and plants. Methods are
CC provided for introducing the retroelements into cells, especially
CC when the retroelement carries at least 1 agronomically-significant
CC characteristic (ACS). In a preferred method, a helper cell line
CC which expresses gag, pol and env sequences is used to enable
CC transfer of a secondary construct which carries an ACS and has
CC retroelement sequences that allow for replication and integration.
CC Claimed isolated nucleic acid molecules comprise a nucleic acid
CC sequence selected from a retroelement primer binding site, envelope,
CC gag, integrase, reverse transcriptase, protease or RNase-H sequence
CC (see AA352254-61). Also provided are plant retroviral particles that
CC particles are used to transfer the nucleic acids into plant cells.
XX
SQ Sequence 4609 BP; 1420 A; 949 C; 1146 G; 1094 T; 0 other;

	Query Match	95.2%;	Score 571.2;	DB 21;	Length 4609;	
	Best Local Similarity	97.0%;	Pred. No. 1.2e-182;			
	Matches	582;	Conservative	0;	Mismatches	18;
					Indels	0;
					Gaps	0;
Oy	1	TTGAGAGGTGGGCATATATTACCCCATCTCTGTACAGACCGTGTGGGTAAAGGCCATGAAGSTG	60			
Dd	1174	TTGGAGGTGGGTCATATTACCCTACTCTGTACAAACGCTTGSGTAAAGCCAGTACAGSTG	1233			
Oy	61	GTTCOCAAGAAAGGTGAATGACAGTGTGTACAAAATGAGAGAAATGACTTGATTAACAACA	120			
Dd	1234	GTTCOCAAGAAAGGTGAATGACAGTGTGTACAAAATGAGAGAAATGACTTGATTAACAACA	1299			
Oy	121	CGAAGTGTCACTGTTGGCGAATGTGTATCGCATATCGCAAGCTGAATGAAGCACACAGG	180			
Dd	1294	CGAACAGTCACTGGCTGGCGAATGTGTATCGCATATCGCAAGCTGAATGAAGCACTACACGG	1353			
Oy	181	AAGGACCAATTTCCCCTTACCTTACCTTAGATGAGATGCTGGAGAGACTTGACGGCAGGCA	240			
Dd	1354	AAGGACCAATTTCCCCTTACCTTACCTTAGATGAGATGCTGGAGAGACTTGACGGCAGGCA	1413			
Oy	241	TACTACTGTTCCTTGAGATGATACCGGGATATCAACAGATCGCGGTAGACCCAGAGAT	300			
Dd	1414	TACTACTGTTCCTTGAGATGATACCGGGATATCAACAGATCGCGGTAGACCCCATAGAT	1473			
Oy	301	CAGGAGAAGACGGCCTTTACATGCCCCCTTGGCGCTTTTGCTTACAGAAAGATGCCATTG	360			
Dd	1474	CAGGAGAAGACGGCTTTACATGCCCCCTTGGCGCTTTTGCTTACAGAAAGATGTCATTG	1533			
Oy	361	GGGTTATGTATGCACACGACCAATTTTCAGAGGTGACATGCTGGCCATTTTTTCAGACATG	420			
Dd	1534	GGGTTATGTATGTATGCACACGACCAATTTTCAGAGGTGACATGCTGGCCATTTTTTCAGACATG	1593			
Oy	421	GTGGAAGAAACATCGAGGTATTTATGAGACACACTCTCGGTTTTTGAGACCCCATTTTAC	480			
Dd	1594	GTGGAAGAAACATCGAGGTATTTATGAGACACACTCTCGGTTTTTGAGACCCCATTTTAC	1655			
Oy	481	AGCTGTTTGAGAAACCTGAGAGGGGTACTTTCAGAGGTGCGAAGAGACATACTGGTACTG	540			
Dd	1654	AGCTGTTTGAGAAACCTGAGAAATGCTACTTCAGAGGTGCGTATGAGACATACTGGTACTG	1713			
Oy	541	AATTGGGAAAAAGTGTCAATTCATGTTGAGAGGGGACATAGTCTTAGGCCACAAAGATTC	600			
Dd	1714	AATTGGGAAAAAGTGTCAATTCATGTTGAGAGGGGACATAGTCTTAGGCCACAAAGATTC	1773			
RESULT 4						
ID	AAZ35280					
xx	AAZ35280 standard; DNA; 597 Bp.					
AC	AAZ35280;					
xx						
DT	27-MAR-2000 (first entry)					

```

XX  Soybean retroelement calypso reverse transcriptase gene.
DE  Retroelement; retrovirus; transgenic plant; gene transfer:
XX  reverse transcriptase gene; calypso; soybean; ss.
XX  Glycine max.
XX  WO9960842-A2.
PN  02-DEC-1999.
XX  28-MAY-1999; 99MO-US11858.
XX  29-MAY-1998; 98US-0087125.
PR  28-MAY-1999; 99US-0087125.
XX  (WRIGHT) WRIGHT D A.
XX  (VOYT/) VOYTAS D F.
PI  Wright DA, Voytas DF;
XX  WPI; 2000-105586/09.
XX  New nucleic acid molecules for imparting agronomically significant
PT  characters to plants, especially soybean
XX  Disclosure; Page 115; 118pp; English.
PS  The present sequence represents the reverse transcriptase gene of
CC  retrovirus-like element (retroelement) calypso of soybean. The
CC  invention provides molecular tools in the form of retroelements and
CC  retroelement-containing vectors, cells and plants. Methods are
CC  provided for introducing the retroelements into cells, especially
CC  when the retroelement carries at least 1 agronomically-significant
CC  characteristic (ACS). In a preferred method, a helper cell line
CC  which expresses gag, pol and env sequences is used to enable
CC  transfer of a secondary construct which carries an ASC and has
CC  retroelement sequences that allow for replication and integration.
CC  claimed isolated nucleic acid molecules comprise a nucleic acid
CC  sequence selected from a retroelement primer binding site, envelope,
CC  gag, integrase, reverse transcriptase, protease or RNase-H sequence
CC  (see AA23524-61). Also provided are plant retroviral particles that
CC  comprise a retrovirus protein encoded by a nucleic acid sequence
CC  encoding a plant retroviral envelope protein and a nucleic acid
CC  including e.g. the present sequence. These plant retroviral
CC  particles are used to transfer the nucleic acids into plant cells.
XX  Sequence 597 BP; 162 A; 121 C; 158 G; 156 T; 0 other;
SQ
Query Match 94.7%; Score 568.2; DB 21; Length 597;
Best Local Similarity 97.0%; Pred. No. 4e-182;
Matches 579; Conservative 0; Mismatches 18; Indels 0; Gaps 0;
QY 1 TTGGAGGCTGGGCTCATATATACCCCATCTCTGACAGCGCTTGGGTAAAGCCCACTACAGTGG 60
DB 1 TTGGAGGCTGGGCTCATATATACCCCATCTCTGACAGCGCTTGGGTAAAGCCCACTACAGTGG 60
QY 61 GTTCCCAAGAAAGGTGGAATGACAGTGTAGAGATGAGAGAAATGACATTTGATACCAACA 120
DB 61 GTTCCCAAGAAAGGTGGAATGACAGTGTAGAGATGAGAGAAATGACATTTGATACCAACA 120
QY 121 CGAAGTGTACTGTGGCGCAATGTGTATCGACTATCCGAAGCTGAATGACCAACGCG 180
DB 121 CGAAGTGTACTGTGGCGCAATGTGTATCGACTATCCGAAGCTGAATGACCAACGCG 180
QY 181 AAGGACATTTCCCTTACCTTTATGATGATGATGATGATGATGATGATGATGATGATGAT 240
DB 181 AAGGACATTTCCCTTACCTTTATGATGATGATGATGATGATGATGATGATGATGATGAT 240
QY 241 TACTACTGTTTCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
DB 241 TACTACTGTTTCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300

```

```

QY 301 CAGAGAGACGCGCTTTATACATGCCCTTGGCGTCTTCTTACAGAGATGCCATTC 360
DB 301 CAGAGAGACGCGCTTTATACATGCCCTTGGCGTCTTCTTACAGAGATGCCATTC 360
QY 361 GGGTATGTATGATCACCAGCCACATTTGAGAGGTGATGCTGCCATTTTTCAGACATG 420
DB 361 GGGTATGTATGATCACCAGCCACATTTGAGAGGTGATGCTGCCATTTTTCAGACATG 420
QY 421 GTGAGAGAAAGCATCGAGGTATTTATGAGAGACTTCTCGGTTTTTGGACCTCATTTGAC 480
DB 421 GTGAGAGAAAGCATCGAGGTATTTATGAGAGACTTCTCGGTTTTTGGACCTCATTTGAC 480
QY 481 AGCTGTTTGAAGAACCTAGAGAGGTACTTTCAGAGGTGCGAAGAGACTTAACCTGTAATG 540
DB 481 AGCTGTTTGAAGAACCTAGAGAGGTACTTTCAGAGGTGCGAAGAGACTTAACCTGTAATG 540
QY 541 AATTGGGAAAAAGTGTCTATTTTCATGCTTTCGAGAGGCGATATGCTTCAAGCCACAAATC 597
DB 541 AATTGGGAAAAAGTGTCTATTTTCATGCTTTCGAGAGGCGATATGCTTCAAGCCACAAATC 597

```

RESULT 5

AA235271 standard; DNA; 9829 BP.

AA235271;

27-MAR-2000 (first entry)

Soybean retroelement Calypso 1-1.

Retroelement; retrovirus; transgenic plant; gene transfer;

Calypso 1-1; soybean; ss.

Glycine max.

WO9960842-A2.

02-DEC-1999.

28-MAY-1999; 99MO-US11858.

29-MAY-1998; 98US-0087125.

28-MAY-1999; 99US-0087125.

(WRIGHT) WRIGHT D A.

(VOYT/) VOYTAS D F.

Wright DA, Voytas DF;

WPI; 2000-105586/09.

New nucleic acid molecules for imparting agronomically significant

characters to plants, especially soybean

Example 3; Page 95-98; 118pp; English.

This is the nucleotide sequence of the Calypso 1-1 retroelement of soybean. It was identified by screening of a soybean lambda library using a reverse transcriptase probe. 2 Groups of soybean retroelements were identified, i.e. Calypso 1-1, 1-2 and 1-3 (see AA235271-73) and Calypso 2-1 and 2-2 (see AA235274-75). The retroelements include gag, pol, env and primer binding site sequences that can be used in constructs of the invention. The invention provides molecular tools in the form of retroelements and retroelement-containing vectors, cells and plants. Methods are provided for introducing the retroelements into cells, especially when the retroelement carries at least 1 agronomically-significant characteristic (ACS). In a preferred method, a helper cell line which expresses gag, pol and env sequences is used to enable transfer of a secondary construct which carries an ASC and has retroelement sequences that allow for replication and integration.

CC Claimed isolated nucleic acid molecules comprise a nucleic acid
 CC sequence selected from a retroelement primer binding site, envelope,
 CC gag, integrase, reverse transcriptase, protease or RNase-H sequence,
 CC (see AA35254-61). Also provided are plant retroviral particles that
 CC particles are used to transfer the nucleic acids into plant cells.

XX Sequence 9829 BP; 2948 A; 2067 C; 2260 G; 2554 T; 0 other;

Query Match 88.7%; Score 532; DB 21; Length 9829;
 Best Local Similarity 94.0%; Pred. No. 3.7e-169;
 Matches 564; Conservative 0; Mismatches 35; Indels 1; Gaps 1;

QY 1 TTGGAGGCTGGGCTCATATACCCATCTCTGACAGCGCTGGGTAAGCCAGTACAGGTG 60
 DB 4271 TTGGAGGCTGGGCTCATATACCCATCTCTGATAGCGCTGGGTAAGTCCAGTACAGGTG 4330
 QY 61 GTTCCCAAGAAAGGTGATGACAGTGTACAGATGAGAGATGACTTGATACCAACA 120
 DB 4331 GTTCCCAAGAAAGCGGATGACAGTGTACAGAAATGAGAGATGACTTGATACCAACA 4390
 QY 121 CGAAGTGTACGTGGTGGGATGTGTATGACTATCGCAAGCTGATGAAGCCACACGG 180
 DB 4391 CGAAGTGTACGTGGTGGGATGTGTATGACTATCGCAAGTGAATGAAGCCACACGG 4450
 QY 181 AAGGACATTTCCCTTACCTTCATGATCAGATGCTGGAGAGACTTGGACGGCAGGCA 240
 DB 4451 AAGGACATTTCCCTTACCTTCATGATCAGATGCTGGAGAGACTTGGACGGCAGGCA 4510
 QY 241 TACTACTGTTTCTTGGATGATATCTCGGATACACCAATCGCGGTAGACCCGAGAGAT 300
 DB 4511 TACTACTGTTTCTTGGATGATATCTCGGATACACCAATCGCGGTAGACCCGAGAGAT 4569
 QY 301 CAGGGAAGACGGCCTTACATGCCCCCTTGGGCTCTTCTTCAACAAGATCCATTC 360
 DB 4570 CAGGGAAGACGGCCTTACATGCCCCCTTGGGCTCTTCTTCAACAAGATCCATTC 4629
 QY 361 GGGTATGTAATGACACCAACATTTAGAGGTGATGCTGGCCATTTTTCAGACATG 420
 DB 4630 GGGTATGTAATGACACCAACATTTAGAGGTGATGCTGGCCATTTTTCAGACATG 4669
 QY 421 GTGGAAGAAAGCATGAGGTATTTATGACGACTTCTCGGTTTGGACCCCTCATTTGAC 480
 DB 4690 GTGGAAGAAAGCATGAGGTATTTATGACGACTTCTCGGTTTGGACCCCTCATTTGAC 4749
 QY 481 AGCTGTTTGAAGAACTTAAGAGGTGATCTTCAAGAGGTGCGCAAGACTTAATGCTACTG 540
 DB 4750 AGCTGTTTGAAGAACTTAAGAGGTGATCTTCAAGAGGTGCGCTAATGACTTAATGCTACTG 4809
 QY 541 AATTGGAAAGATGTCATTTGATGTTGAGAGGGGCTAGTCTAGGCCACAAGATCTCA 600
 DB 4810 AATTGGAAAGATGTCATTTGATGTTGAGAGGGGCTAGTCTAGGCCACAAGATCTCA 4869

RESULT 6
 AA35272
 ID AA35272 standard; DNA; 12571 BP.

XX AA35272;
 XX 27-MAR-2000 (first entry)
 XX Soybean retroelement Calypso 1-2.
 DE Retroelement; retrovirus; transgenic plant; gene transfer;
 KM Calypso 1-2; soybean; ss.
 XX Glycine max.
 OS WO960842-A2.
 PM 02-DEC-1999.
 PD 28-MAY-1999; 99WO-US11858.
 XX

XX 29-MAY-1998; 98US-0087125.
 PR 28-MAY-1999; 99US-0087125.
 XX
 XX (WRIG/) WRIGHT D A.
 PA (VOYT/) VOYTAS D F.
 XX
 PI Wright DA, Voytas DF.
 XX WPI: 2000-105586/09.
 DR
 XX
 PT New nucleic acid molecules for imparting agronomically significant
 characters to plants, especially soybean

Example 3; Page 98-102; 118pp; English.

This is the nucleotide sequence of the Calypso 1-2 retroelement of soybean. It was identified by screening of a soybean lambda library using a reverse transcriptase probe. 2 groups of soybean retroelements were identified, i.e. Calypso 1-1, 1-2 and 1-3 (see AA35271-73) and Calypso 2-1 and 2-2 (see AA35274-75). The retroelements include gag, pol, env and primer binding site sequences that can be used in constructs of the invention. The invention provides molecular tools in the form of retroelements and retroelement-containing vectors, cells and plants. Methods are provided for introducing the retroelements into cells, especially when the retroelement carries at least 1 agronomically-significant characteristic (ACS). In a preferred method, a helper cell line which expresses gag, pol and env sequences is used to enable transfer of a secondary construct which carries an ASC and has retroelement sequences that allow for replication and integration. CC Claimed isolated nucleic acid molecules comprise a nucleic acid sequence selected from a retroelement primer binding site, envelope, gag, integrase, reverse transcriptase, protease or RNase-H sequence (see AA35254-61). Also provided are plant retroviral particles that particles are used to transfer the nucleic acids into plant cells.

Sequence 12571 BP; 3801 A; 2382 C; 2707 G; 3661 T; 0 other;

Query Match 87.2%; Score 523.2; DB 21; Length 12571;
 Best Local Similarity 93.2%; Pred. No. 4.1e-166;
 Matches 559; Conservative 0; Mismatches 38; Indels 3; Gaps 1;

QY 1 TTGGAGGCTGGGCTCATATACCCATCTCTGACAGCGCTGGGTAAGCCAGTACAGGTG 60
 DB 4295 TTGGAGGCTGGGCTCATATACCCATCTCTGACAGCGCTGGGTAAGCCAGTACAGGTG 4354
 QY 61 GTTCCCAAGAAAGGTGATGACAGTGTACAGATGAGAGATGACTTGATACCAACA 120
 DB 4355 GTTCCCAAGAAAGGTGATGACAGTGTACAGATGAGAGATGACTTGATACCAACA 4414
 QY 121 CGAAGTGTACGTGGTGGGATGTGTATGACTATGCAAGCTGAAAGCCACACGG 180
 DB 4415 CGAAGTGTACGTGGTGGGATGTGTATGACTATGCAAGCTGAAAGCCACACGG 4474
 QY 181 AAGGACATTTCCCTTACCTTCATGATCAGATGCTGGAGAGACTTGCAGGGCAGGCA 240
 DB 4475 AAGGACATTTCCCTTACCTTCATGATCAGATGCTGGAGAGACTTGCAGGGCAGGCA 4554
 QY 241 TACTACTGTTTCTTGGATGATATCTCGGATACACCAATCGCGGTAGACCCAGAT 300
 DB 4535 TACTACTGTTTCTTGGATGATATCTCGGATATATATCAAGTCCGGTGGACCCAGAT 4594
 QY 301 CAGGGAAGACGGCCTTACATGCCCCCTTGGGCTCTTCTTCAACAAGAGATCCATTC 360
 DB 4595 CAGGGAAGACGGCCTTACATGCCCCCTTGGGCTCTTCTTCAACAAGAGATCCATTC 4654
 QY 361 GGGTATGTAATGACACCAACATTTAGAGGTGATGCTGGCCATTTTTCAGACATG 420
 DB 4655 GGGTATGTAATGACACCAACATTTAGAGGTGATGCTGGCCATTTTTCAGACATG 4714
 QY 421 GTGGAAGAAAGCATGAGGTATTTATGACGACTTCTCGGTTTGGACCCCTCATTTGAC 480

PT New nucleic acid molecules for imparting agronomically significant
 PT characters to plants, especially soybean -
 XX
 PS Claim 36(a); Page 113; 118pp; English.

CC The present sequence represents the reverse transcriptase gene of
CC retrovirus-like element (retroelement) cyclops of pea. The
CC invention provides molecular tools in the form of retroelements and
CC retroelement-containing vectors, cells and plants. Methods are
CC provided for introducing the retroelements into cells, especially
CC when the retroelement carries at least 1 agronomically-significant
CC characteristic (ASC). In a preferred method, a helper cell line
CC which expresses gag, pol and env sequences is used to enable
CC transfer of a secondary construct which carries an ASC and has
CC retroelement sequences that allow for replication and integration.
CC Claimed isolated nucleic acid molecules comprise a nucleic acid
CC sequence selected from a retroelement primer binding site, envelope,
CC gag, integrase, reverse transcriptase, protease or RNase-H sequence
CC (see AA35254-61). Also provided are plant retroviral particles that
CC comprise a retrovirus protein encoded by a nucleic acid sequence
CC encoding a plant retroviral envelope protein and a nucleic acid
CC including e.g. the present sequence. These plant retroviral
CC particles are used to transfer the nucleic acids into plant cells.

SQ Sequence 597 BP; 158 A; 121 C; 152 G; 166 T; 0 other;

Query Match	52.3%	Score 313.8	DB 21	Length 597
Best Local Similarity	70.4%	Pred. No. 8.4e-96		
Matches 420; Conservative	0	Mismatches 177	Indels 0	Gaps 0

QY	1	TTGGAGCGTGGGCATCATATACCACCTCTGCAGACGGCTTGGGTAAAGCCAGTAAAGTG	60
Db	1	TTGGATCGGAGAAATGATTTTACCAGATCTCGGATATGCTCATGGGTACGTCCTGATGTG	60
QY	61	GTTCGGAAGAAAGGTGGAAATACCGTGATCCGGATGAGCAAGATGATTTGATTCCTTACC	120
Db	61	GTTCGGAAGAAAGGTGGAAATACCGTGATCCGGATGAGCAAGATGATTTGATTCCTTACC	120
QY	121	CGAACTGCTACTGTTGGCGAATGTGATGCATATCGCAAGCTGATGAAGCCACACGG	180
Db	121	AAAGTTCCAAACGGGGTGGAGAAATGTATTAATATAGCGCGTTGAATACCGCAACTCGA	180
QY	181	AAGGACCAATTTCCCTTAACCTTTATGATGATCAGATGCTGGAGAGACTTCGAGGGACGA	240
Db	181	AAGGACCAATTTCCACATCCCTTATGATGATTAATGCTGGAAGAGCTCTCCGGGCAACA	240
QY	241	TACTACTGTTTCTTGATGATCTCGGATACCAACGATCGCGGTAGACCCCGAGAT	300
Db	241	TACTATTTGTTCTTGATGATGCTATTCGGGGATATACCAAAATTTGCCGTTGACCCGGCGAT	300
QY	301	CAGGAGAAGACGGCCTTTACATGCCCTTTGGCGCTTTTGCTTTACAGAAGATGCCATTTC	360
Db	301	CATTAAAAAGACGGCTTTACATGCTCCGTTTGAGATGTTGCGATACCGAAAAATGTCCTTT	360
QY	361	GGGTATGATGATGACACGACCAATTCAGAGGGATCGTGGCCATTTTTCAGACATG	420
Db	361	GGGTATGTCATGACACCGACGACATTTTCCAAAGATGTGTGAAGCCATTTTTCGCACTT	420
QY	421	GTGGAGAAAAGCATCGAGGTATTTATGAGCACATCTTCGTTTTGACCCCTCATTTTAC	480
Db	421	AATAGAGAAAACAAATGGAAGCTTCATGAGTACATCTTCGTGATTTTGGGTATTCCTTAA	480
QY	481	AGCTGTTGAGGAACCTTAGAGAGGGTACTCAGAGGTGCGAAGACCTTAACCTTGGTACTG	540
Db	481	TTATGCTTTGGCAAACTTGAAAACGCTGCTTGAAGAATGTGGAAGACCAATCTTGTCTT	540
QY	541	AATTTGGGAAAAGTCAATTCATGTTGATGATGATGAGGCAATATGCTTAAGGCCAACAATG	597
Db	541	AATTTGTTGAAGTCCCACTTCATGTTGATGATGATGAGGCAATATGCTTGGCCATAAATG	597

RESULT 9
AAF22302/C

ID	AAF22302 standard; DNA; 94895 BP.
1	

AC AAF22302;

DT 20-MAR-2001 (first entry)

DE BAC containing repeats from centromeres 1-4 #25.

KW Centromere; michromosome; vector; ds

OS *Arabidopsis thaliana*.

PN W0200055325-A2.

PD 21-SEP-2000.

PF 17-MAR-2000; 2000WO-US07392.

PR 18-MAR-1999; 99US-0125219.

PR 18-MAY-1999; 99US-0134770.

PR 17-SEP-1999; 99US-0154603.

PA (UYCH-) UNIV CHICAGO.

PI Preuss D, Copenhagen G, Ke

DR WPI; 2000-587529/55.

PT Recombinant DNA construct comprising

PT the construction of

PS Claim 102; Page 804-819; 1449pp; English.

CC The present invention relates to a recombinant DNA construct of a plant

CC producing stably inherited microsomes which can serve as vectors for

CC selected proteins such as hormones, enzymes, interleukins, clotting

XX

Current: M3+Q4
E1 38
C33330 307 8
DE 21
T032+P 0190E1

Best Local Similarity 69.68; Pred. No. 1.5e-92;

QY 1 TTGGAGGCTGGGCTCATTATACCCCATCTGTGCACAGCGCTGGGTAAGCCCACTACAGTGG 60
DB 35132 TTAGAGTTGGGGTTATTACTACCCATCTGTGATAGCACTTGSGGTATCTCCACTTCATTGCG 35072
QY 61 GTTCCCAGAAGAGGTGGGAATGACAGTGTACGAGATGAGAGAAATGACTTGTATCCAMCA 120
DB 35072 GTTCCATAAAGAGAGGTATGACAGTGTAAAAATCTTAAGATGAACATGATACCACACT 35011
QY 121 CGAAGCTGCACGTGGTGGCCAATGTATGCACTATGCGAAAGCTGAATGAAACCCACACGG 180
DB 35012 AAGACATATACATGACAATAGAGTTGTGATTGACATATAGAAAGTTAAAAGTCTCCCTYTACA 34953
QY 181 AAGGACCATTTCCCTTACCTTCATGAGATGAGATGCGTAGAGACTTGCAGGGCAGGCA 240
DB 34952 AAAGACCATTTTCCCATTTGCCCTTCATTTGATGATACGATACTAGAAAGTTATAGCAACCATCT 34899
QY 241 TACTACTGTTTTCTTGATGGATACACTCGGGATACAACAGATCGGGTAGACCCCAAGAT 300
DB 34892 TACTATTGCTTTTGTATGGTATTAGCGGATTTTTTCAAATCCCATTTACACCAAAATGAC 34833
QY 301 CAGGAGAAAGCGGCTTTTACATGCCCCCTTGGCGCTTTGGTTACAGAAAGATGGCATTC 360
DB 34832 CAGAGAAAAAACCACTTTCACCTTGCTCCATTATGGAGACTTGGCTTACAAAGCGTATGCCCTTTC 34773

```
QY 361 GGGTATATGATGACACACCATTTTCAGAGGTCATGCTGGCCATTTTTCAGACATG 420
  || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 34772 GGTCTATATGACACACCATTTTTCAGAGGTCATGCTGGCCATTTTTCAGACATG 34713
QY 421 GTGGAGAAAGCATGAGGTATTTATGACGACCTTCGTTTTCAGACCTCATTTTAC 480
  || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 34712 ATAGAGGAGATGATGAGGTATTCATGATGATTTTCTGTATGCTGCTCTCTCTC 34653
QY 481 AGCTTTTTCAGAGGAGGTATTCATGAGGTATTCATGAGGTATTCATGAGGTATTC 540
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 34652 TCGTCTTTTTCAGAGGTATTCATGAGGTATTCATGAGGTATTCATGAGGTATTC 34593
QY 541 AATTGGGAAAGTGTCTATTTTCATGAGGTATTCATGAGGTATTCATGAGGTATTC 599
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 34592 AACTGGGAGAAATGCCATTTTCATGAGGTATTCATGAGGTATTCATGAGGTATTC 34534

RESULT 10
AAF22305
ID AAF22305 standard; DNA: 1082138 BP.
AC AAF22305;
XX
XX 20-MAR-2001 (first entry)
DT
XX
XX Arabidopsis thaliana chromosome 4 centromere.
DE
XX
XX Centromere; mitochondrion; vector; ds.
KM
XX
XX Arabidopsis thaliana.
OS
XX
XX WO200055325-A2.
PN
XX
XX 21-SEP-2000.
PD
XX
XX 17-MAR-2000; 2000WO-US07392.
PE
XX
XX 18-MAR-1999; 99US-0125219.
PR 01-APR-1999; 99US-0127409.
PR 18-MAY-1999; 99US-0134770.
PR 13-SEP-1999; 99US-0135584.
PR 17-SEP-1999; 99US-0154603.
XX
XX (UYCH-) UNIV CHICAGO.
PA
XX
XX Preus D, Copenhagen G, Keith K;
PI
XX
XX WPI; 2000-587529/55.
DR
XX
XX Recombinant DNA construct comprising a plant centromere, useful for
PT producing stably inherited mitochondria which can serve as vectors for
PT the construction of transgenic plant and animal cells
XX
XX Claim 68; Page 977-1388; 1449pp; English.
PS
XX
XX The present invention relates to a recombinant DNA construct of a plant
CC (Arabidopsis thaliana) centromere. The constructs are useful for
CC producing stably inherited mitochondria which can serve as vectors for
CC the construction of transgenic plant and animal cells expressing
CC selected proteins such as hormones, enzymes, interleukins, clotting
CC factors, cytokines, antibodies, and growth factors.
XX
XX
XX Sequence 1082138 BP; 348775 A; 194404 C; 195515 G; 343444 T; 0 other;
SQ

Query Match 51.38; Score 307.8; DB 21; Length 1082138;
Best Local Similarity 69.68; Pred. No. 5.9e-92;
Matches 417; Conservative 0; Mismatches 182; Indels 0; Gaps 0;
```

```
Db 1047067 GTCTTAAAGAGGATGATGACATGTTTAAAGATGATGATGATGATGATGATGAT 1047126
QY 121 CGAAGTGTACATGTTTTCAGAGGTATTCATGAGGTATTCATGAGGTATTCATGAGGTATTC 180
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1047127 AGGATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1047186
QY 181 AAGAGCATTTTCCCTTACCTTTCATGAGGTATTCATGAGGTATTCATGAGGTATTC 240
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1047187 AAGAGCATTTTCCCTTACCTTTCATGAGGTATTCATGAGGTATTCATGAGGTATTC 1047246
QY 241 TACTAGTGTCTTTCATGAGGTATTCATGAGGTATTCATGAGGTATTCATGAGGTATTC 300
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1047247 TACTAGTGTCTTTCATGAGGTATTCATGAGGTATTCATGAGGTATTCATGAGGTATTC 1047306
QY 301 CAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1047307 CAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1047366
QY 361 GGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 420
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1047367 GGTCTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1047426
QY 421 GTGGAGAAAGCATGAGGTATTTATGAGGTATTCATGAGGTATTCATGAGGTATTCATG 480
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1047427 ATAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1047486
QY 481 AGCTGTTTTCAGAGGTATTCATGAGGTATTCATGAGGTATTCATGAGGTATTCATGAG 540
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1047487 TCGTGTGTTTTCAGAGGTATTCATGAGGTATTCATGAGGTATTCATGAGGTATTC 1047546
QY 541 AATTGGGAAAGTGTCTATTTTCATGAGGTATTCATGAGGTATTCATGAGGTATTCATG 599
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1047547 AACTGGGAGAAATGCCATTTTCATGAGGTATTCATGAGGTATTCATGAGGTATTC 1047605

RESULT 11
AA235264
ID AA235264 standard; DNA: 600 BP.
AC AA235264;
XX
XX 27-MAR-2000 (first entry)
DT
XX
XX Arabidopsis retroelement athila reverse transcriptase gene.
DE
XX
XX Retroelement; retrovirus; transgenic plant; gene transfer;
KM reverse transcriptase gene; athila; ss.
KW
XX
XX Arabidopsis thaliana.
OS
XX
XX WO9960842-A2.
PN
XX
XX 02-DEC-1999.
PD
XX
XX 28-MAY-1999; 99US-011858.
PR 29-MAY-1999; 99US-0087125.
PR 28-MAY-1999; 99US-0087125.
XX
XX (WRIGHT) WRIGHT D A.
PA (VOIT/) VOIT D F.
XX
XX Wright DA, Voytas DF;
PI
XX
XX WPI; 2000-105586/09.
DR
XX
XX New nucleic acid molecules for imparting agronomically significant
PT characters to plants, especially soybean
PT
XX
XX Claim 36(a); Page 112-113; 118pp; English.
PS
XX
XX The present sequence represents the reverse transcriptase gene of
CC
```


CC producing stably inherited microsomes which can serve as vectors for
 CC the construction of transgenic plant and animal cells expressing
 CC selected proteins such as hormones, enzymes, interleukins, clotting
 CC factors, cytokines, antibodies, and growth factors.

XX Sequence 90336 BP; 26524 A; 18837 C; 18625 G; 26350 T; 0 other;

Query Match 50.5%; Score 303; DB 21; Length 90336;
 Best Local Similarity 69.1%; Pred. No. 6.3e-91;
 Matches 414; Conservative 0; Mismatches 185; Indels 0; Gaps 0;

OY 1 TTGGAGCGTGGGCTCATATATACCCCATCTCTGACAGCGCTTGGTAAAGCCGATACAGGTG 60
 DB 58215 TTAGATGCGGGGTTATTTACCTATCTCTGATGAGCACTGGGTATCTCCAGTTCACTGC 58274
 OY 61 GTTCCCAAGAAAGTGGCAATGACAGTGTGATGAGATGAGAGATGATGATACCAACA 120
 DB 58275 GTTCCCAAGAAAGTGGCAATGACAGTGTGATGAGATGAGAGATGATGATACCAACT 58334
 OY 121 CGAAGTGTACGTGGTGGGATGTGTATCGATATCGCAAGCTGATGAAAGCCACACGG 180
 DB 58335 AGGACTATATCTCGACATAGAGATGTATGACTATAGGAAGTAAATGCTGCTCTAGA 58394
 OY 181 AAGACCAATTTCCCTTACCTTATGATGATGATGATGATGATGATGATGATGATGATG 240
 DB 58395 AAGACCAATTTCCCTTACCTTATGATGATGATGATGATGATGATGATGATGATGATG 58454
 OY 241 TACTACTGTCTTCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 300
 DB 58455 TACTACTGTCTTCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 58514
 OY 301 CAGGAGAGAGCGCTTACATGCCCCCTTGGCGTCTTCTTACAGAAAGATGCCATTC 360
 DB 58515 CAGGAGAGAGCGCTTACATGCCCCCTTGGCGTCTTCTTACAGAAAGATGCCATTC 58574
 OY 361 GGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 420
 DB 58575 GGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 58634
 OY 421 GTGGAGAAAGCATGAGATATTTATGAGACTTCCGTTTGGACCTCATTTGAC 480
 DB 58635 ATATGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 58694
 OY 481 AGCTTTTGGAGACCTAGAGAGGTACTTCAAGAGTGCAGAGAGACTTAAGTGGTACTG 540
 DB 58695 TCGTGTCTTGAACCTATGAAAGGTACTTCAAAAGATGTCAGAGACAAACCTGCTG 58754
 OY 541 AATGGGAAAGTGTCAATTTGATGTTGCGAGAGGCAATGCTTAGGCCACAAAGATCTC 599
 DB 58755 AACTGGAGAAATGCAATTTGATGTTGCGAGAGGCAATGCTTAGGCCACAAATTTTC 58813

RESULT 15

ID AAF22302 standard; DNA; 94895 BP.

AC AAF22302;

DT 20-MAR-2001 (first entry)

DE BAC containing repeats from centromeres 1-4 #25.

XX Centromere; michrosome; vector; ds.

OS Arabidopsis thaliana.

PN WO20005325-A2.

PD 21-SEP-2000.

PF 17-MAR-2000; 2000WO-US07392.

PR 18-MAR-1999; 99US-0125219.

PR 01-APR-1999; 99US-0127409.
 PR 18-MAR-1999; 99US-0134770.
 PR 13-SEP-1999; 99US-013584.
 PR 17-SEP-1999; 99US-0154603.

XX (UYCH-) UNIV CHICAGO.

PI Preuss D, Copenhagen G, Keith K;

DR WPI; 2000-587529/55.

XX Recombinant DNA construct comprising a plant centromere, useful for
 PT producing stably inherited microsomes which can serve as vectors for
 PT the construction of transgenic plant and animal cells

PS Claim 102; Page 804-819; 1449pp; English.

CC The present invention relates to a recombinant DNA construct of a plant
 CC (Arabidopsis thaliana) centromere. The constructs are useful for
 CC producing stably inherited microsomes which can serve as vectors for
 CC the construction of transgenic plant and animal cells expressing
 CC selected proteins such as hormones, enzymes, interleukins, clotting
 CC factors, cytokines, antibodies, and growth factors.

XX Sequence 94895 BP; 28943 A; 18101 C; 18466 G; 29385 T; 0 other;

Query Match 50.5%; Score 303; DB 21; Length 94895;
 Best Local Similarity 69.1%; Pred. No. 6.5e-91;
 Matches 414; Conservative 0; Mismatches 185; Indels 0; Gaps 0;

OY 2 TGGAGCGTGGGCTCATATATACCCCATCTCTGACAGCGCTTGGTAAAGCCGATACAGGTG 61
 DB 45866 TTGATGTTGGTGTACATCTATCCCTATCTGATGATGATGATGATGATGATGATGATG 45925
 OY 62 TTCCCAAGAAAGTGGCAATGACAGTGTGATGAGATGAGAGATGATGATGATGATGATG 121
 DB 45926 TTCCCAAGAAAGTGGCAATGACAGTGTGATGAGATGAGAGATGATGATGATGATG 45985
 OY 122 GAACGTGTACGTGGTGGCAATGTGTATGACTATGCAAGCTGATGAAAGCCACACGGA 181
 DB 45986 GAACGTGTACGTGGTGGCAATGTGTATGACTATGCAAGCTGATGAAAGCTGATG 46045
 OY 182 AGACCAATTTCCCTTACCTTATGATGATGATGATGATGATGATGATGATGATGATG 241
 DB 46046 AAGATCATTTTCTTACCTTATGATGATGATGATGATGATGATGATGATGATGATG 46105
 OY 242 ACTACTGTTCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 301
 DB 46106 ACTACTGTTCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 46165
 OY 302 AGGAGAAAGCGGCTTTACATGCCCCCTTGGCGTCTTCTTACAGAGAGATGCCATTCG 361
 DB 46166 AAGAGAAAGCGGCTTTACATGCCCCCTTGGCGTCTTCTTACAGAGAGATGCCATTCG 46225
 OY 362 GGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 421
 DB 46226 GTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 46285
 OY 422 TGGAGAAAGCATGAGATATTTATGAGACTTCCGTTTGGACCTCATTTTGAACCTTGA 481
 DB 46286 TGGAGAGATGATGAGAGGTCTTCAAGAGATGATGATGATGATGATGATGATGATG 46345
 OY 482 GCTGTTGAGGAACCTAGAGAGGCTCTTACAGAGTGCAGAGACTTAATCTTGGTACTGA 541
 DB 46346 CATGTTGTTGAGTGTGAGAGGTCTTATGAGAGGCTGTGAGAGCAAACTGTTCTCA 46405
 OY 542 ATTGGGAAAGTGTCAATTTGATGTTGCGAGAGGCAATGCTTAGGCCACAAAGATCTC 600
 DB 46406 ATTGGGAAAGTGTCAATTTGATGTTGCGAGAGGCAATGCTTAGGCCACAAAGATCTC 46464

Search completed: June 20, 2003, 21:58:33
 Job time : 345.396 secs

[illegible]

Query Match	100.0%	Score 600	DB 4	Length 600
Best Local Similarity	100.0%	Pred. No. 6	4e-199	
Matches	600	Conservative 0	Mismatches 0	Indels 0
			Gaps 0	
QY	1	TTGGAGCGTGGCCTCATATACCCCATCTCTGTGACAGCGCTTGGGTAAAGCCAGTACAGGTG	60	
Db	1	TTGGAGCGTGGCCTCATATACCCCATCTCTGTGACAGCGCTTGGGTAAAGCCAGTACAGGTG	60	
QY	61	GTGCCAAGAAGGTGGGATGACATGTGTACAGAGATGAGAGAAATGACTGTATACCACAA	120	
Db	61	GTGCCAAGAAGGTGGGATGACATGTGTACAGAGATGAGAGAAATGACTGTATACCACAA	120	
QY	121	CGAAGTCGACTGGTGTGGCGAATGTGTATCGACTATCGCAAGCTGATATGACCCACACGG	180	
Db	121	CGAAGTCGACTGGTGTGGCGAATGTGTATCGACTATCGCAAGCTGATATGACCCACACGG	180	
QY	181	AAGGACCATTTCCCTTACCTTTTCATGTGATTCAGATGCTGTGAGAGACTTGCAGGACGCA	240	
Db	181	AAGGACCATTTCCCTTACCTTTTCATGTGATTCAGATGCTGTGAGAGACTTGCAGGACGCA	240	
QY	241	TACTACTGTTCCTTGATGATGATACACGGGATTCACAACAGATGCGCGTAGACCCACAGAGAT	300	
Db	241	TACTACTGTTCCTTGATGATGATACACGGGATTCACAACAGATGCGCGTAGACCCACAGAGAT	300	
QY	301	CAGAGAAAGACGGCCTTTACATGCCCTTTGGCGTCTTGGCTTACAGAAAGATGCCATTC	360	
Db	301	CAGAGAAAGACGGCCTTTACATGCCCTTTGGCGTCTTGGCTTACAGAAAGATGCCATTC	360	

Db 301 CAGGAGAAAGCGCCCTTTACATGCCCTTTGGCGTCTTGTCTTACAGAAAGATGCCATTTC 360
QY 361 GGGTTATGTAATGACACCGACCATTTTCAGAGTGCATCTGCGCATTTTTCACAGATG 420
Db 361 GGGTTATGTAATGACACCGACCATTTTCAGAGTGCATCTGCGCATTTTTCACAGATG 420
QY 421 GTGGAGAAAAGCATCGAGGATTTATGACGACTTCTGGTTTGGACCCCTCATTTTTCAG 480
Db 421 GTGGAGAAAAGCATCGAGGATTTATGACGACTTCTGGTTTGGACCCCTCATTTTTCAG 480
QY 481 AGCGTTTGGAGAACTAGAGAGGTACTTTCAGAGTGCAGAGTGCAGACTTTCAGAGTGC 540
Db 481 AGCGTTTGGAGAACTAGAGAGGTACTTTCAGAGTGCAGAGTGCAGACTTTCAGAGTGC 540
QY 541 AATTGGAAAAAGTGTCAATTTTCATGTTTCGAGAGGCGATAGTCTTAGGCCACAAATCTCA 600
Db 541 AATTGGAAAAAGTGTCAATTTTCATGTTTCGAGAGGCGATAGTCTTAGGCCACAAATCTCA 600

RESULT 2
US-09-322-478-17
; Sequence 17, Application US/09322478
; Patent No. 6331662
; GENERAL INFORMATION:
; APPLICANT: Wright, David A.
; APPLICANT: Voytas, Daniel F.
; TITLE OF INVENTION: Plant Retroelements and Methods Related Thereto
; FILE REFERENCE: P-1065 ISURF Plant Retroelement
; CURRENT APPLICATION NUMBER: US/09/322.478
; EARLIER FILING DATE: 1998-05-28
; EARLIER APPLICATION NUMBER: 60/087125
; EARLIER FILING DATE: 1998-05-29
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 17
; LENGTH: 12286
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: plant
; OTHER INFORMATION: retroelement sequence
US-09-322-478-17

Query Match 100.0%; Score 600; DB 4; Length 12286;
Best Local Similarity 100.0%; Pred. No. 3.8e-198;
Matches 600; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGGAGGCTGGGCTCATATATCCCATCTCTGACAGCCGCTTGGGTAAAGCCCATACAGTG 60
Db 4278 TTGGAGGCTGGGCTCATATATCCCATCTCTGACAGCCGCTTGGGTAAAGCCCATACAGTG 4337
QY 61 GTTCCCAAGAAAGGTGGAATGACAGTGGTACAGATGAGAGAAATGACTTGTATACCA 120
Db 4338 GTTCCCAAGAAAGGTGGAATGACAGTGGTACAGATGAGAGAAATGACTTGTATACCA 4397
QY 121 CGAAGTGTCACTGGTGGCGAATGTATGCACTATGCGAAGCTGAATGAACCCACACGG 180
Db 4398 CGAAGTGTCACTGGTGGCGAATGTATGCACTATGCGAAGCTGAATGAACCCACACGG 4457
QY 181 AAGGACCATTTCCCTTACCTTTCATGATGATGATGATGATGATGATGATGATGATGATG 240
Db 4458 AAGGACCATTTCCCTTACCTTTCATGATGATGATGATGATGATGATGATGATGATGATG 4517
QY 241 TACTACTGTTTCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 300
Db 4518 TACTACTGTTTCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4577
QY 301 CAGGAGAAAGCGGCTTTTACATGCGCTTTGGCGTCTTGTCTTACAGAAAGATGCCATTTC 360
Db 4578 CAGGAGAAAGCGGCTTTTACATGCGCTTTGGCGTCTTGTCTTACAGAAAGATGCCATTTC 4637
QY 361 GGGTTATGTAATGACACCGACCATTTTCAGAGTGCATCTGCGCATTTTTCACAGATG 420

Db 4638 GGGTTATGTAATGACACCGACCATTTTCAGAGTGCATCTGCGCATTTTTCACAGATG 4697
QY 421 GTGGAGAAAAGCATCGAGGATTTATGACGACTTCTGGTTTGGACCCCTCATTTTTCAG 480
Db 4698 GTGGAGAAAAGCATCGAGGATTTATGACGACTTCTGGTTTGGACCCCTCATTTTTCAG 4757
QY 481 AGCGTTTGGAGAACTAGAGAGGTACTTTCAGAGTGCAGAGTGCAGACTTTCAGAGTGC 540
Db 4758 AGCGTTTGGAGAACTAGAGAGGTACTTTCAGAGTGCAGAGTGCAGACTTTCAGAGTGC 4817
QY 541 AATTGGAAAAAGTGTCAATTTTCATGTTTCGAGAGGCGATAGTCTTAGGCCACAAATCTCA 600
Db 4818 AATTGGAAAAAGTGTCAATTTTCATGTTTCGAGAGGCGATAGTCTTAGGCCACAAATCTCA 4877

RESULT 3
US-09-322-478-21
; Sequence 21, Application US/09322478
; Patent No. 6331662
; GENERAL INFORMATION:
; APPLICANT: Wright, David A.
; APPLICANT: Voytas, Daniel F.
; TITLE OF INVENTION: Plant Retroelements and Methods Related Thereto
; FILE REFERENCE: P-1065 ISURF Plant Retroelement
; CURRENT APPLICATION NUMBER: US/09/322.478
; EARLIER FILING DATE: 1998-05-28
; EARLIER APPLICATION NUMBER: 60/087125
; EARLIER FILING DATE: 1998-05-29
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 21
; LENGTH: 4609
; TYPE: DNA
; ORGANISM: Glycine max
US-09-322-478-21

Query Match 95.2%; Score 571.2; DB 4; Length 4609;
Best Local Similarity 97.0%; Pred. No. 2.2e-188;
Matches 582; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 1 TTGGAGGCTGGGCTCATATATCCCATCTCTGACAGCCGCTTGGGTAAAGCCCATACAGTG 60
Db 1174 TTGGAGGCTGGGCTCATATATCCCATCTCTGACAGCCGCTTGGGTAAAGCCCATACAGTG 1233
QY 61 GTTCCCAAGAAAGGTGGAATGACAGTGGTACAGATGAGAGAAATGACTTGTATACCA 120
Db 1234 GTTCCCAAGAAAGGTGGAATGACAGTGGTACAGATGAGAGAAATGACTTGTATACCA 1293
QY 121 CGAAGTGTCACTGGTGGCGAATGTATGCACTATGCGAAGCTGAATGAACCCACACGG 180
Db 1294 CGAAGTGTCACTGGTGGCGAATGTATGCACTATGCGAAGCTGAATGAACCCACACGG 1353
QY 181 AAGGACCATTTCCCTTACCTTTCATGATGATGATGATGATGATGATGATGATGATGATG 240
Db 1354 AAGGACCATTTCCCTTACCTTTCATGATGATGATGATGATGATGATGATGATGATGATG 1413
QY 241 TACTACTGTTTCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 300
Db 1414 TACTACTGTTTCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1473
QY 301 CAGGAGAAAGCGGCTTTTACATGCGCTTTGGCGTCTTGTCTTACAGAAAGATGCCATTTC 360
Db 1474 CAGGAGAAAGCGGCTTTTACATGCGCTTTGGCGTCTTGTCTTACAGAAAGATGCCATTTC 1533
QY 361 GGGTTATGTAATGACACCGACCATTTTCAGAGTGCATCTGCGCATTTTTCACAGATG 420
Db 1534 GGGTTATGTAATGACACCGACCATTTTCAGAGTGCATCTGCGCATTTTTCACAGATG 1593
QY 421 GTGGAGAAAAGCATCGAGGATTTTATGACGACTTCTGGTTTGGACCCCTCATTTTTCAG 480
Db 1594 GTGGAGAAAAGCATCGAGGATTTTATGACGACTTCTGGTTTGGACCCCTCATTTTTCAG 1653
QY 481 AGCGTTTGGAGAACTAGAGAGGTACTTTCAGAGTGCAGAGTGCAGACTTTCAGAGTGC 540

Db 1624 AGCTGTTGAGAACCTAGAAATGTTACTTCAGAGGTCGGTAGAGACTAATGTTGTTACTG 1713
QY 541 AATTGGGAAAGTGCATTTTCATGTTGTCAGAGGGGATAGTCTTAGGCCCAAGATCTCA 600
Db 1714 AATTGGGAAAGTGCATTTTATGTTGTCAGAGGGGATAGTCTTAGGCCCAAGATCTCA 1773

RESULT 4

US-09-322-478-34
; Sequence 34, Application US/09322478
; Patent No. 6331662
; GENERAL INFORMATION:
; APPLICANT: Wright, David A.
; APPLICANT: Voytas, Daniel F.
; TITLE OF INVENTION: Plant Retroelements and Methods Related Thereto
; FILE REFERENCE: P-1065 ISURF Plant Retroelement
; CURRENT APPLICATION NUMBER: US/09/322,478
; EARLIER FILING DATE: 1999-05-28
; EARLIER APPLICATION NUMBER: 60/087125
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 34
; LENGTH: 597
; TYPE: DNA
; ORGANISM: Glycine max
US-09-322-478-34

Query Match 94.7%; Score 568.2; DB 4; Length 597;
Best Local Similarity 97.0%; Pred. No. 7.5e-188;
Matches 579; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 1 TTGAGGCTGGGCTATATATACCCCATCTCTGACAGCGCTTGGTAAAGCCAGTACAGGTG 60
Db 1 TTGAGGCTGGGCTATATATACCCCATCTCTGACAGCGCTTGGTAAAGCCAGTACAGGTG 60
QY 61 GTTCCCAAGAAAGTGGATGACAGTGTGATGACATGACAGTGTGATGACAGTGTGATGACAG 120
Db 61 GTTCCCAAGAAAGTGGATGACAGTGTGATGACATGACAGTGTGATGACAGTGTGATGACAG 120
QY 121 CGAAGTGTCTGCTGTTGGCGAATGTGTATGACATGACAGTGTGATGACAGTGTGATGACAG 180
Db 121 CGAAGTGTCTGCTGTTGGCGAATGTGTATGACATGACAGTGTGATGACAGTGTGATGACAG 180
QY 181 AAGGACCTTTCCCTTACCTTTCATGATGATGATGATGATGATGATGATGATGATGATGAT 240
Db 181 AAGGACCTTTCCCTTACCTTTCATGATGATGATGATGATGATGATGATGATGATGATGAT 240
QY 241 TACTACTGTTCTTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
Db 241 TACTACTGTTCTTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
QY 301 CAGGAGAGAGCGGCTTTACATGCCCTTTGGCGTCTTGGCTTTACAGAAAGATGCCATGTC 360
Db 301 CAGGAGAGAGCGGCTTTACATGCCCTTTGGCGTCTTGGCTTTACAGAAAGATGCCATGTC 360
QY 361 GGGTATATATGACACCGCCATTTTACAGAGTGTGATGATGATGATGATGATGATGATGATGAT 420
Db 361 GGGTATATATGACACCGCCATTTTACAGAGTGTGATGATGATGATGATGATGATGATGATGAT 420
QY 421 GTGGAGAAAGCATGAGATTTTATGACAGCTTCTCGGTTTGGAGCCCATTTTATGAC 480
Db 421 GTGGAGAAAGCATGAGATTTTATGACAGCTTCTCGGTTTGGAGCCCATTTTATGAC 480
QY 481 AGCTGTTTGGAGAACCTAGAAATGTTACTTCAGAGGTCGGTAGAGACTAATGTTGTTACTG 540
Db 481 AGCTGTTTGGAGAACCTAGAAATGTTACTTCAGAGGTCGGTAGAGACTAATGTTGTTACTG 540
QY 541 AATTGGGAAAGTGCATTTTCATGTTGTTGTCAGAGGGGATAGTCTTAGGCCCAAGATCTCA 597
Db 541 AATTGGGAAAGTGCATTTTATGTTGTTGTCAGAGGGGATAGTCTTAGGCCCAAGATCTCA 597

RESULT 5

US-09-322-478-19
; Sequence 19, Application US/09322478
; Patent No. 6331662
; GENERAL INFORMATION:
; APPLICANT: Wright, David A.
; APPLICANT: Voytas, Daniel F.
; TITLE OF INVENTION: Plant Retroelements and Methods Related Thereto
; FILE REFERENCE: P-1065 ISURF Plant Retroelement
; CURRENT APPLICATION NUMBER: US/09/322,478
; EARLIER FILING DATE: 1999-05-28
; EARLIER APPLICATION NUMBER: 60/087125
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 19
; LENGTH: 9829
; TYPE: DNA
; ORGANISM: Glycine max
US-09-322-478-19

Query Match 88.7%; Score 532; DB 4; Length 9829;
Best Local Similarity 94.0%; Pred. No. 1.5e-174;
Matches 564; Conservative 0; Mismatches 35; Indels 1; Gaps 1;

QY 1 TTGAGGCTGGGCTATATATACCCCATCTCTGACAGCGCTTGGTAAAGCCAGTACAGGTG 60
Db 4271 TTGAGGCTGGGCTATATATACCCCATCTCTGACAGCGCTTGGTAAAGCCAGTACAGGTG 4330
QY 61 GTTCCCAAGAAAGTGGATGACAGTGTGATGACATGACAGTGTGATGACAGTGTGATGACAG 120
Db 4331 GTTCCCAAGAAAGTGGATGACAGTGTGATGACATGACAGTGTGATGACAGTGTGATGACAG 4390
QY 121 CGAAGTGTCTGCTGTTGGCGAATGTGTATGACATGACAGTGTGATGACAGTGTGATGACAG 180
Db 4391 CGAAGTGTCTGCTGTTGGCGAATGTGTATGACATGACAGTGTGATGACAGTGTGATGACAG 4450
QY 181 AAGGACCTTTCCCTTACCTTTCATGATGATGATGATGATGATGATGATGATGATGATGAT 240
Db 4451 AAGGACCTTTCCCTTACCTTTCATGATGATGATGATGATGATGATGATGATGATGATGAT 240
QY 241 TACTACTGTTCTTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
Db 4511 TACTACTGTTCTTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4569
QY 301 CAGGAGAGAGCGGCTTTACATGCCCTTTGGCGTCTTGGCTTTACAGAAAGATGCCATGTC 360
Db 4570 CAGGAGAGAGCGGCTTTACATGCCCTTTGGCGTCTTGGCTTTACAGAAAGATGCCATGTC 4629
QY 361 GGGTATATATGACACCGCCATTTTACAGAGTGTGATGATGATGATGATGATGATGATGATGAT 420
Db 4630 GGGTATATATGACACCGCCATTTTACAGAGTGTGATGATGATGATGATGATGATGATGATGAT 4689
QY 421 GTGGAGAAAGCATGAGATTTTATGACAGCTTCTCGGTTTGGAGCCCATTTTATGAC 480
Db 4690 GTGGAGAAAGCATGAGATTTTATGACAGCTTCTCGGTTTGGAGCCCATTTTATGAC 4749
QY 481 AGCTGTTTGGAGAACCTAGAAATGTTACTTCAGAGGTCGGTAGAGACTAATGTTGTTACTG 540
Db 4750 AACCTATTTGGAGAACCTAGAAATGTTACTTCAGAGGTCGGTAGAGACTAATGTTGTTACTG 4809
QY 541 AATTGGGAAAGTGCATTTTCATGTTGTTGTCAGAGGGGATAGTCTTAGGCCCAAGATCTCA 600
Db 4810 AATTGGGAAAGTGCATTTTATGTTGTTGTCAGAGGGGATAGTCTTAGGCCCAAGATCTCA 4869

RESULT 6

US-09-322-478-20
; Sequence 20, Application US/09322478
; Patent No. 6331662
; GENERAL INFORMATION:
; APPLICANT: Wright, David A.

```

; APPLICANT: Voytas, Daniel F.
; TITLE OF INVENTION: Plant Retroelements and Methods Related Thereto
; FILE REFERENCE: P-1065 ISURF Plant Retroelement
; CURRENT APPLICATION NUMBER: US/09/322,478
; CURRENT FILING DATE: 1999-05-28
; EARLIER APPLICATION NUMBER: 60/087125
; EARLIER FILING DATE: 1998-05-29
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
; LENGTH: 12571
; TYPE: DNA
; ORGANISM: Glycine max
US-09-322-478-20

Query Match      87.2%; Score 523.2; DB 4; Length 12571;
Best Local Similarity 93.2%; Pred.No. 2.1e-171;
Matches 559; Conservative 0; Mismatches 38; Indels 3; Gaps 1;

OY 1 TTGGAGGCTGGGCTCATATACCCCTCTCTGACAGCGCTTGGGTAGCCAGTACAGTG 60
DB 4295 TTGGAGGCTGGGCTCATATACCCCTCTCTGACAGCGCTTGGGTAGCCAGTACAGTG 4354
OY 61 GTTCCCAAGAAAGGTGGAATGACAGTGTGATGAGATGAGAGAAATGACTTGATCCACA 120
DB 4355 GTTCCCAAGAAAGGTGGAATGACAGTGTGATGAGAAATGAGAAATGACTTGATCCACA 4414
OY 121 CGAAGCTGCACTGTGGGCAATGTGTATGCACTGTGCAAGCTGTAAGTAAACCAACGG 180
DB 4415 CGAAGCTGCACTGTGGGCAATGTGTATGCACTGTGCAAGCTGTAAGTAAACCAACGA 4474
OY 181 AAGGACCATTTCCCTTACCTTTCATGATGATGATGATGATGATGATGATGATGATG 240
DB 4475 AAGGACCATTTCCCTTACCTTTCATGATGATGATGATGATGATGATGATGATGATG 4534
OY 241 TACTACTGTTTCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 300
DB 4535 TACTACTGTTTCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4594
OY 301 CAGGAGAAAGCGGCTTTACATGCGCTTGGCGCTTGGCTTACAGAAAGATGCAATTC 360
DB 4595 CAGGAGAAAGCGGCTTTACATGCGCTTGGCGCTTGGCTTACAGAAAGATGCAATTC 4654
OY 361 GGGTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 420
DB 4655 GGGTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4714
OY 421 GTGAGAGAAAGCATGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 480
DB 4715 GTGAGAGAAAGCATGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 4774
OY 481 AGCTGTTTGAAGAACTGAGAGAGGCTTCTGAGAGGTGCGAAGAGACTTACTGTTAGT 540
DB 4775 A--GTTTGAAGAACTGAGAGAGGCTTCTGAGAGGTGCGAAGAGACTTACTGTTAGT 4831
OY 541 AATTGGGAAAAGTGTCTATTTCTGATGTTGAGAGGGGCTATGCTTGAAGCCACAGATCTCA 600
DB 4832 AACTGGGAAAAGTGTCTATTTCTGATGTTGAGAGGGGCTATGCTTGAAGCCACAGATCTCA 4891

RESULT 7
US-09-322-478-22
; Sequence 22, Application US/09322478
; Patent No. 6331662
; GENERAL INFORMATION:
; APPLICANT: Wright, David A.
; APPLICANT: Voytas, Daniel F.
; TITLE OF INVENTION: Plant Retroelements and Methods Related Thereto
; FILE REFERENCE: P-1065 ISURF Plant Retroelement
; CURRENT APPLICATION NUMBER: US/09/322,478
; CURRENT FILING DATE: 1999-05-28
; EARLIER APPLICATION NUMBER: 60/087125
; EARLIER FILING DATE: 1998-05-29
;
; ORGANISM: P1sum sativum
US-09-322-478-29

```

```

; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 22
; LENGTH: 9139
; TYPE: DNA
; ORGANISM: Glycine max
US-09-322-478-22

Query Match      57.6%; Score 345.6; DB 4; Length 9139;
Best Local Similarity 73.5%; Pred.No. 1.1e-109;
Matches 441; Conservative 0; Mismatches 159; Indels 0; Gaps 0;

OY 1 TTGGAGGCTGGGCTCATATACCCCTCTCTGACAGCGCTTGGGTAGCCAGTACAGTG 60
DB 1646 TAGGAGCAGGCGCTTATTTACCCCTCGGATGATGATGATGATGATGATGATGATGATG 1705
OY 61 GTTCCCAAGAAAGGTGGAATGACAGTGTGATGAGATGAGAGAAATGACTTGATCCACA 120
DB 1706 GTTCCCAAGAAAGGTGGAATGACAGTGTGATGAGAAATGAGAAATGACTTGATCCACA 1765
OY 121 CGAAGCTGCACTGTGGGCAATGTGTATGCACTGTGCAAGCTGTAAGTAAACCAACGG 180
DB 1766 AGGACTGTCAACCGGCTGAGAAATGTGATGATGATGATGATGATGATGATGATGATG 1825
OY 181 AAGGACCATTTCCCTTACCTTTCATGATGATGATGATGATGATGATGATGATGATGATG 240
DB 1826 AAGGACCATTTCCCTTACCTTTCATGATGATGATGATGATGATGATGATGATGATGATG 1885
OY 241 TACTACTGTTTCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 300
DB 1886 TATTATGTTTCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1945
OY 301 CAGGAGAAAGCGGCTTTACATGCGCTTGGCGCTTGGCTTACAGAAAGATGCAATTC 360
DB 1946 CAGGAGAAAGCGGCTTTACATGCGCTTGGCGCTTGGCTTACAGAAAGATGCAATTC 2005
OY 361 GGGTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 420
DB 2006 GGGTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2065
OY 421 GTGAGAGAAAGCATGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 480
DB 2066 GTGAGAGAAAGCATGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 2125
OY 481 AGCTGTTTGAAGAACTGAGAGAGGCTTCTGAGAGGTGCGAAGAGACTTACTGTTAGT 540
DB 2126 GGGTGGCTTATTAATCTTGAAGAGATTTACAGAGATGTGAAGATGCCAATCTAGTCTC 2185
OY 541 AATTGGGAAAAGTGTCTATTTCTGATGTTGAGAGGGGCTATGCTTGAAGCCACAGATCTCA 600
DB 2186 AATTGGGAAAAGTGTCTATTTCTGATGTTGAGAGGGGCTATGCTTGAAGCCACAGATCTCA 2245

RESULT 8
US-09-322-478-29
; Sequence 29, Application US/09322478
; Patent No. 6331662
; GENERAL INFORMATION:
; APPLICANT: Wright, David A.
; APPLICANT: Voytas, Daniel F.
; TITLE OF INVENTION: Plant Retroelements and Methods Related Thereto
; FILE REFERENCE: P-1065 ISURF Plant Retroelement
; CURRENT APPLICATION NUMBER: US/09/322,478
; CURRENT FILING DATE: 1999-05-28
; EARLIER APPLICATION NUMBER: 60/087125
; EARLIER FILING DATE: 1998-05-29
;
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 29
; LENGTH: 597
; TYPE: DNA
; ORGANISM: P1sum sativum
US-09-322-478-29

```


Accession	Sequence	Position
Db	GAATGTCATGATTATTCGGAAGTTGAAATATATGCCACTTCGAAAGACCAATTATCCACTCC	4183
QY	CTTTCATGGAATCAATGCTGGAGAGACTTGCAGGGCAGGCATTAAGTCTTTCTGGATG	259
Db	CTTTCATGGAACCAATGCTTGAGAGAGACTCGCAAGGCATCATATTATTTGTTCTGGATG	4243
QY	GATACCTGGGATACACCAAGATCCCGGTAGACCCGAGATCAGAGA	305
Db	GATTTCTAGTTACAAATTAGATTCTTATAGCATATCAAGATCAAGA	4289

RESULT 11
US-08-855-449-2/c
; Sequence 2, Application US/08855449
; Date of US/08855449

GENERAL INFORMATION:
APPLICANT: AKAMATSU, TOKYOKAZU
APPLICANT: SUZUKI, TAKAO
TITLE OF INVENTION: METHOD FOR IDENTIFYING THE SEX OF
TITLE OF INVENTION: SPINACH BY DNA MARKERS
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
ADDRESS: P.C.
STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/855,449
FILING DATE: 13-MAY-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 119124/1996
FILING DATE: 14-MAY-1996
ATTORNEY/AGENT INFORMATION:
NAME: OBLON, NORMAN F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 7828-0003-0
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-413-3000
TELEFAX: 703-413-2220
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1347 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE: SPINACH
ORGANISM: SPINACH

Query Match	11.58;	Score 69;	DB 2;	Length 1347;
Best Local Similarity	51.18;	Pred. No. 6.6e-14;		
Matches 162; Conservative	0;	Mismatches 155;	Indels 0;	Gaps 0;

QY	139	CGAATGATGATACGACTATCCGACAGTGAATGAAGCCACACGAGGAGACATTTCCCTCTTA	198
Db	1111	CAATGTGCGCCGATTACAGATATCTTATACACGACTACGCCCTTAAGACGATTTTCCACTCG	105
QY	199	CCCTTACATGATACAGATCCTGGAAGACTTCCAGGCGACATACACTACTGTTTCTTGAT	258
Db	1051	CGACTATATCCGATCTTGTGTCAGACACACACCAATTAATGCTTACTCTCTTTTATGTAAC	992
QY	259	GGATCTCTGGGATTCACACACAGATCGCGGTGAGCCCCAGAGATTCAGAGAGAGACGCGCCCTTT	318

D _b	991	TGTTACGAGCGCTACAATGCAGTCCCATAGCAGAGGACATGGAGAAGCAACCCTTC	932
O _y	319	ACATGCCCCCTTTGGSGCTCTTCTGTACAGAAAGATGCCATTGGCGTTATGTATGCACCA	378
D _b	931	ATCACCTTAAGTGGGGTGACATATGTCTAACAGTAGTCCGTTGGACTGGAAGAACACGGGG	872
O _y	379	GCCACATTTCAGAGGTGCATGCTGCCAATTTTTTCAGACATGCTGGAGAAAAGCATCGAG	438
D _b	871	GCTACCTCTCAAATTAACAACCAACACCATCATAGGACATGATTAAAGAGAAATTGAG	812
O _y	439	GTATTTATGAGCAGACTT	455
D _b	811	GAATAATGTCAGCAGCAT	795

RESULT 12
US-08-694-869-2
; Sequence 2, Application US/08694865
; Patent No. 5004132

APPLICANT: Olszewski, N.
APPLICANT: Tzafir, I.
APPLICANT: Somers, D. A.
APPLICANT: Lockhart, B.
APPLICANT: Torbert, K.
TITLE OF INVENTION: SUGARCANE BACILLIFORM VIRUS
TITLE OF INVENTION: PROMOTER
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Schwegman, Lundberg, Woessner & Kluth, P.A.
STREET: P.O. Box 2938
CITY: Minneapolis
STATE: MN
COUNTRY: USA
ZIP: 55402
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/694,869
FILING DATE: 09-AUG-1996
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Woessner, Warren D
REGISTRATION NUMBER: 30,440
REFERENCE/DOCKET NUMBER: 600.369US1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612-373-6900
TELEFAX: 612-339-3061
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 7566 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
S-08-694-869-2

Query Match	2	Score	55.4	DB	2	Length	7568
Best Local Similarity	46.6%	Pred	9.8e-09				
Matches	213	Conservative	0	Mismatches	241	Indels	3
				Gaps			1

140	GAATGTGTCACACTATCGCAAGCTGAATGACGCCACACGGAAGACATTTCCCTTAC	199
5482	GACTGGTTTCATTATTCACAGAGGCTAAAGTACACACATGCGCGATCATTAATTCATTGC	5541

QY 200 CTTTCATGATCAGATGCTGGAGAGACTTGCAGGCGCAGCACTACTGTTCTTGATG 259
 Db 5542 CCGGATCAATGCTGCTACTATAAAATGTTGCAAGAGCAAAAGATCTTCAAGTTGAT 5601
 QY 260 GATCTCGGGATACACAGATGCGGGTAGACCCAGAGATGAGAGAAAGCGCCTTA 319
 Db 5602 TGAAGAGGGGTTTCATCAAGCGCCATGATGAGAAAGTATTCATTAAAGCATTTT 5661
 QY 320 CATGCCCTTTGGGCTTTTGGTTACAGAGATGCCATTCGGCTATGTAAGCCAG 379
 Db 5662 CAGATACATGATGCTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5721
 QY 380 CCACATTCAGAGGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 439
 Db 5722 CAATCTCCAGAGAAAATGAGCCAGTGTTCAGAGAAACAGAGGG--TTCATAGCTG 5778
 QY 440 TATTATGAGAGACTCTCGGTTTGGACCTCATTTGACAGCTTTTGAGAACCTAG 499
 Db 5779 TGTATATGATGATGATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5838
 QY 500 AGAGGACTCTCAGAGGTGCGAAGAGACTAACTGCTGCTGCTGCTGCTGCTGCT 559
 Db 5839 GGAAGATGCTTCAATCTGCAAAAGAGATGCTGCTGCTGCTGCTGCTGCTGCT 5898
 QY 560 TCATGCTGAGAGGGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 596
 Db 5899 TAGAGTTAAGAGGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5935

RESULT 13

US-09-349-546-2
 ; Sequence 2, Application US/09349546
 ; Patent No. 6093569
 ; GENERAL INFORMATION:
 ; APPLICANT: Olszewski, N.
 ; APPLICANT: Tzafir, I.
 ; APPLICANT: Somers, D.A.
 ; APPLICANT: Lockhart, B.
 ; APPLICANT: Torbert, K.
 ; TITLE OF INVENTION: Sugarcane bacilliform virus promoter
 ; FILE REFERENCE: 600.369US2
 ; CURRENT APPLICATION NUMBER: US/09/349,546
 ; EARLIER FILING DATE: 1999-07-08
 ; EARLIER APPLICATION NUMBER: US 08/694,869
 ; EARLIER FILING DATE: 1996-08-09
 ; EARLIER APPLICATION NUMBER: PCT/IB97/01338
 ; EARLIER FILING DATE: 1997-08-13
 ; NUMBER OF SEQ ID NOS: 10
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 2
 ; LENGTH: 7568
 ; TYPE: DNA
 ; ORGANISM: sugarcane bacilliform virus
 ; US-09-349-546-2

Query Match 9.2%; Score 55.4; DB 3; Length 7568;
 Best Local Similarity 46.6%; Pred. No. 9, 8e-09;
 Matches 213; Conservative 0; Mismatches 241; Indels 3; Gaps 1;
 QY 140 GAATGTATGACTATTCGCAAGCTGAATGAAGCCACAGAGACCAATTCCTCCCTTAC 199
 Db 5482 GACTGGTTTCAATTAACAGAGGCTAATGACACACAGATGCGGATCAATATTCATG 5541
 QY 200 CTTTCATGATCAGATGCTGGAGAGACTTGCAGGCGCAGCATACACTGTTCTTGATG 259
 Db 5542 CCGGATCAATGCTGCTACTATAAAATGTTGCAAGAGCAAAAGATCTTCAAGTTGAT 5601
 QY 260 GATCTCGGGATACACAGATGCGGGTAGACCCAGAGATGAGAGAAAGCGCCTTA 319
 Db 5602 TGAAGAGGGGTTTCATCAAGCGCCATGATGAGAAAGTATTCATTAAAGCATTTT 5661
 QY 320 CATGCCCTTTGGGCTTTTGGTTACAGAGATGCCATTCGGCTATGTAAGCCAG 379
 Db 5779 TGTATATGATGATGATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5838

Db 5662 CAGATACATGAGCTGATGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5721
 QY 380 CCACATTCAGAGGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 439
 Db 5722 CAATCTCCAGAGAAAATGAGCCAGTGTTCAGAGAAACAGAGGG--TTCATAGCTG 5778
 QY 440 TATTATGAGAGACTCTCGGTTTGGACCTCATTTGACAGCTTTTGAGAACCTAG 499
 Db 5779 TGTATATGATGATGATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5838
 QY 500 AGAGGACTCTCAGAGGTGCGAAGAGACTAACTGCTGCTGCTGCTGCTGCTGCT 559
 Db 5839 GGAAGATGCTTCAATCTGCAAAAGAGATGCTGCTGCTGCTGCTGCTGCTGCT 5898
 QY 560 TCATGCTGAGAGGGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 596
 Db 5899 TAGAGTTAAGAGGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5935

RESULT 14

US-09-360-186-1/c
 ; Sequence 1, Application US/09360186
 ; Patent No. 6262343
 ; GENERAL INFORMATION:
 ; APPLICANT: Staskawicz, et al.
 ; TITLE OF INVENTION: Bsz Resistance Gene
 ; FILE REFERENCE: 50687
 ; CURRENT APPLICATION NUMBER: US/09/360,186
 ; EARLIER FILING DATE: 1999-07-23
 ; EARLIER APPLICATION NUMBER: 60/093,957
 ; EARLIER FILING DATE: 1998-07-23
 ; NUMBER OF SEQ ID NOS: 9
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 1
 ; LENGTH: 31491
 ; TYPE: DNA
 ; ORGANISM: Capsicum annuum
 ; US-09-360-186-1

Query Match 8.1%; Score 48.6; DB 4; Length 31491;
 Best Local Similarity 46.8%; Pred. No. 5, 3e-06;
 Matches 153; Conservative 0; Mismatches 174; Indels 0; Gaps 0;
 QY 139 CGAATGTATGACTATTCGCAAGCTGAATGAAGCCACAGAGACCAATTCCTCCCTTA 198
 Db 24307 CCGATGCTGATGACTTACCGTCAAGTGAATGAATGATCATTAATAATATCTCTT 24248
 QY 199 CTTTCATGATCAGATGCTGGAGAGACTTGCAGGCGCAGCATACTGTTCTTGAT 258
 Db 24247 CCGATGCTGATGACTTTCGACAGCTTCAAGGCTGTAAGTCTTTCACAAATATGAC 24188
 QY 259 GGATACCTGGGATACACAGATGCGGGTAGACCCAGAGATGAGAGAAAGCGCCTT 318
 Db 24187 CTTGCTGCGGTTTACATCAAGTGAATAATAGAGATGACATACCCAAAGAGCGCTTC 24128
 QY 319 ACATGCCCTTTGGGCTTTGCTTACAGAGATGCCATTCGGGTTATGTAAGCCACCA 378
 Db 24127 CGAACCCATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 24068
 QY 379 GCCAATTTTCAGAGGTGATGCTGCGCATTTTTCAGAGATGGTGAAGAAAGCATCGAG 438
 Db 24067 GCAGCTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 24008
 QY 439 GTATTATGAGAGACTCTCGGTTT 465
 Db 24007 GTATTATGATTAATATCTGATCTAT 23981
 RESULT 15
 US-09-392-054-5
 ; Sequence 5, Application US/09592054
 ; Patent No. 6440684
 ; GENERAL INFORMATION:

APPLICANT: Berand, Christophe
APPLICANT: Finer, Jeffrey
APPLICANT: Sakowicz, Roman
APPLICANT: Wood, Kenneth
TITLE OF INVENTION: No. 6440684el motor proteins and methods for
FILE REFERENCE: 1016
CURRENT APPLICATION NUMBER: US/09/592.054
CURRENT FILING DATE: 2000-07-20
NUMBER OF SEQ ID NOS: 8
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5
LENGTH: 1421
TYPE: DNA
ORGANISM: Human
US-09-592-054-5

Query Match 5.6%; Score 33.8; DB 4; Length 1421;
Best Local Similarity 51.7%; Pred. No. 0.12;
Matches 77; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 53 TACAGGTGTTCCCAAGAAAGCTGATGACAGTGTGAGAGTGAAGAGTGA 112
DB 484 TACGAGAGATCCTTAAGAGGACATTAAGATTGGGACTCAGTGAAGAGACTGTTTG 543
QY 113 TACCAACACGAACTGTCAGTGTGGCAATGTGATGCACTATGCAAGCTGAATGAG 172
DB 544 TTGCTTGTGATAGTGTCTCTGTTGGACAGGCGCACTAGAGTGTGGCTTCA 603
QY 173 CCACACGGAAGAACATTTCCCTTACCT 201
DB 604 CGGCTATGAAGTCCAGTCCGCGATCT 632

Search completed: June 20, 2003, 23:25:48
Job time : 76.6415 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2003, 21:58:41 ; Search time 221.698 Seconds
(without alignments)
3971.420 Million cell updates/sec

Title: US-09-965-553-11

Perfect score: 600

Sequence: 1 ttggagctgtgcatcata.....tcttagccacaagatctca 600

Scoring table:

IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1042519 seqs, 733713590 residues

Total number of hits satisfying chosen parameters: 2085038

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published_Applications_NA.*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	600	100.0	600	10	US-09-965-553-11 Sequence 11, Appl
2	600	100.0	12286	10	US-09-965-553-17 Sequence 17, Appl
3	571.2	95.2	4609	10	US-09-965-553-21 Sequence 21, Appl
4	568.2	94.7	597	10	US-09-965-553-34 Sequence 34, Appl
5	532	88.7	9829	10	US-09-965-553-19 Sequence 19, Appl
6	523.2	87.2	12571	10	US-09-965-553-20 Sequence 20, Appl
7	345.6	57.6	9139	10	US-09-965-553-22 Sequence 22, Appl
8	313.8	52.3	597	10	US-09-965-553-29 Sequence 29, Appl
9	304.6	50.8	600	10	US-09-965-553-27 Sequence 27, Appl
10	172.4	28.7	10482	10	US-09-965-553-23 Sequence 23, Appl
11	80.8	11.5	3813	9	US-10-083-357-634 Sequence 634, App
12	68.4	11.4	631	9	US-10-255-536-16 Sequence 16, Appl
13	64	10.7	4298	9	US-09-997-672-1 Sequence 1, Appl1
14	64	10.7	4921	9	US-09-997-672-2 Sequence 2, Appl1
15	48.6	8.1	31491	10	US-09-864-680-1 Sequence 1, Appl1
16	38.8	6.5	642	9	US-10-123-155-370 Sequence 370, App
17	34.8	5.8	541	9	US-10-184-644-152 Sequence 152, App
18	34.8	5.8	541	9	US-10-184-634-152 Sequence 152, App
19	33.6	5.6	369	10	US-09-960-352-4371 Sequence 4371, App

c	20	33.4	5.6	1355	9	US-10-021-577-1	Sequence 1, Appl1
	21	33	5.5	360	9	US-10-012-896-438	Sequence 438, App
	22	33	5.5	360	9	US-09-895-793-438	Sequence 438, App
	23	33	5.5	360	9	US-09-895-814-438	Sequence 438, App
	24	33	5.5	360	9	US-10-010-940-438	Sequence 438, App
	25	33	5.5	360	10	US-09-759-143-438	Sequence 438, App
	26	33	5.5	360	10	US-09-780-669-438	Sequence 438, App
	27	33	5.5	360	10	US-09-822-827-438	Sequence 438, App
	28	33	5.5	377	10	US-09-960-352-9257	Sequence 9257, App
	29	33	5.5	378	10	US-09-960-352-12818	Sequence 12818, A
	30	33	5.5	379	10	US-09-960-352-9011	Sequence 9011, App
c	31	33	5.5	408	10	US-09-960-352-10425	Sequence 10425, A
	32	33	5.5	425	10	US-09-960-352-2591	Sequence 2591, App
c	33	33	5.5	470	9	US-09-918-995-8924	Sequence 8924, App
	34	33	5.5	2187	9	US-10-175-523-910	Sequence 110, App
c	35	31.4	5.2	500	10	US-09-925-300-825	Sequence 825, App
	36	31.4	5.2	671	9	US-10-184-644-346	Sequence 346, App
c	37	31.4	5.2	1584	10	US-09-732-224-3	Sequence 346, App
c	38	31	5.2	1629	10	US-09-732-224-6	Sequence 3, Appl1
c	39	31	5.2	1629	10	US-09-732-224-6	Sequence 6, Appl1
c	40	30.8	5.1	809	9	US-09-764-891-2253	Sequence 2253, App
c	41	30.8	5.1	809	9	US-10-205-428-181	Sequence 181, App
c	42	30.6	5.1	472	9	US-09-918-995-33903	Sequence 33903, A
c	43	30.6	5.1	699	9	US-10-184-644-138	Sequence 138, App
c	44	30.6	5.1	699	9	US-10-184-634-138	Sequence 138, App
c	45	30.6	5.1	6271	10	US-09-993-292A-1	Sequence 1, Appl1

ALIGNMENTS

RESULT 1
US-09-965-553-11
Sequence 11, Application US/09965553
Patent No. US20020112259A1
GENERAL INFORMATION:
APPLICANT: Wright, David A.
TITLE OF INVENTION: Plant Retroelements and Methods Related Thereto
FILE REFERENCE: P-1065 ISURF Plant Retroelement
CURRENT APPLICATION NUMBER: US/09/965, 553
PRIOR FILING DATE: 2001-09-27
PRIOR APPLICATION NUMBER: 09/322, 478
PRIOR FILING DATE: 1999-05-28
PRIOR APPLICATION NUMBER: 60/087125
NUMBER OF SEQ ID NOS: 41
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 11
LENGTH: 600
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: plant
US-09-965-553-11

Query Match 100.0%; Score 600; DB 10; Length 600;
Best Local Similarity 100.0%; Pred. No. 4.9e-197;
Matches 600; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	TTGGAGGCGGGCTCATATACCCATCTGACAGCGCTGGGTAGCCAGTACAGCTG 60
DB	1	TTGGAGGCTGGGCTCATATACCCATCTGACAGCGCTGGGTAGCCAGTACAGCTG 60
QY	61	GTCCCAAGAAAGGTGAGTACAGTGTGACAGAGAGAGATGACTGATACACACA 120
DB	61	GTCCCAAGAAAGGTGAGTACAGTGTGACAGAGAGAGATGACTGATACACACA 120
QY	121	CGAATGTGCTGCTGTTGGCGAATGTGTATGACATACGCAAGTGAATGACACACG 180
DB	121	CGAATGTGCTGCTGTTGGCGAATGTGTATGACATACGCAAGTGAATGACACACG 180

OY	181	AAGGACCACTTTCCTTACCTTTCATGATACAGATCTGTGAGAGACTTGCAGGGCAGGCA	240
Db	181	AAGGACCACTTTCCTTACCTTTCATGATACAGATCTGTGAGAGACTTGCAGGGCAGGCA	240
OY	241	TACTACTGTTTCTTGGATGATACACGGGATACAAACACATCGCGGTAGACCCACAGAT	300
Db	241	TACTACTGTTTCTTGGATGATACACGGGATACAAACACATCGCGGTAGACCCACAGAT	300
OY	301	CAGGAGAAGACGGCCTTTACATGCCCTTTGGCGTCTTGCCTTACAGAAGGATGCCATTG	360
Db	301	CAGGAGAAGACGGCCTTTACATGCCCTTTGGCGTCTTGCCTTACAGAAGGATGCCATTG	360
OY	361	GGGTATGTATGATGACACACGCCACATTTCAAGGTGATCCTGTGCCATTTTTTCAGACATG	420
Db	361	GGGTATGTATGATGACACACGCCACATTTCAAGGTGATCCTGTGCCATTTTTTCAGACATG	420
OY	421	GTGGAGAAAGCATGAGGTATTTATGACGACCTTCGTGGTTTTTGACCCTCATTTGAC	480
Db	421	GTGGAGAAAGCATGAGGTATTTATGACGACCTTCGTGGTTTTTGACCCTCATTTGAC	480
OY	481	AGCTGTTTGAGGCAACCTAGAGAGGGGTACTTCAAGAGTGGGAAGAGACAATCTGGTACAG	540
Db	481	AGCTGTTTGAGGCAACCTAGAGAGGGGTACTTCAAGAGTGGGAAGAGACAATCTGGTACAG	540
OY	541	AATTTGGGAAAAGTGTCAATTTCTATGTTGAGAGAGGCAATGTCCTTAGGCCACCAAGATCTA	600
Db	541	AATTTGGGAAAAGTGTCAATTTCTATGTTGAGAGAGGCAATGTCCTTAGGCCACCAAGATCTA	600

```

RESULT 2
US-09-965-553-17
; Sequence 17, Application US/09965553
; Patent No. US20020112259A1
GENERAL INFORMATION:
APPLICANT: Wright, David A.
APPLICANT: Voytas, Daniel F.
TITLE OF INVENTION: Plant Retroelements and Methods Related Thereto
FILE REFERENCE: P-1065 ISURF Plant Retroelement
CURRENT APPLICATION NUMBER: US/09/965,553
CURRENT FILING DATE: 2001-09-27
PRIOR APPLICATION NUMBER: 09/322,478
PRIOR FILING DATE: 1999-05-28
PRIOR APPLICATION NUMBER: 60/087125
PRIOR FILING DATE: 1998-05-29
NUMBER OF SEQ ID NOS: 41
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 17
LENGTH: 12286
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: plant
US-09-965-553-17

```

Query Match	Similarity	100.0%	Score 600	DA 10	Length 12286
Best Local	Similarity	100.0%	Pred. No. 2.3e16		
Matches	600	Conservative	0	Mismatches	0
				Indels	Gaps
QY	1	TTGAGGCGTGGGCTGCATATATACCCCATCTCTGTGACAGCGGCTTGGGTAAGCCCACTAGAGGTG	60		
Db	4278	TTGGAGGGCTGGGCTCATATATACCCCATCTCTGTGACAGCGGCTTGGGTAAGCCCACTAGAGGTG	433		
QY	61	GTTGCCAAGAAGGTGTAATGACATGTGTAGCAGATGAGAGAAATGACTTGTATACCAACA	120		
Db	4338	GTTCCCAAGAAGGTGTAATGACATGTGTAGCAGATGAGAGAAATGACTTGTATACCAACA	439		
QY	121	CGAAGTGTGATGGTGTGGCGAATGTGTATCGACTATGCCAAGCTGAATGAAAGCCACACGG	180		
Db	4398	CGAAGTGTGATGGTGTGGCGAATGTGTATCGACTATGCCAAGCTGAATGAAAGCCACACGG	445		
QY	181	AAGGACCATTTCCCTTACCTTTTACGTGATCAGATGCTGGAGAGACTTGACAGGACAGCA	240		

Db	4458	AAGGACCATTTCCCTTACCTTTTCATGAGATCGTGAGAGACTTGACGGGACGCA	4517
Qy	241	TACTACTGTTCTTGATGATGATACTCGGGATACACAGATCGCGGTAGACCCAGAGAT	300
Db	4518	TACTACTGTTCTTGATGATGATGATACTCGGGATACACAGATCGCGGTAGACCCAGAGAT	4577
Qy	301	CAGGAGAAAGCGGCTTTACATGCCCCCTTTGGCGCTTTTGCTTACAGAGAGATGCCATTTC	360
Db	4578	CAGGAGAGACGGGCTTTACATGCCCTTTGGCGCTTTTGCTTACAGAGAGATGCCATTTC	4637
Qy	361	GGGTATGTATATGACCCAGCCACATTTTACAGAGGTGATGCTGGCCATTTTTCACAGATG	420
Db	4638	GGGTATGTATATGACCCAGCCACATTTTACAGAGGTGATGCTGGCCATTTTTCACAGATG	4697
Qy	421	GTGAGAAACATGAGGATATTTATGTGACACATCTCTGGTTTTTGGACCCCTCATTTTAC	480
Db	4698	GTGAGAAACATGAGGATATTTATGTGACACATCTCTGGTTTTTGGACCCCTCATTTTAC	4757
Qy	481	AGCTGTTTGAAGAACCTTAGAGAGGGTACTTCAGAGGTGCGAAGAGACTTAACTTGTTACTG	540
Db	4758	AGCTGTTTGAAGAACCTTAGAGAGGGTACTTCAGAGGTGCGAAGAGACTTAACTTGTTACTG	4817
Qy	541	AATGCGAAAAGTGTCACTTTCATGTTTGAGAGAGGCAATGTCCTTAAGGCCAACAAGATCTCA	600
Db	4818	AATGCGAAAAGTGTCACTTTCATGTTTGAGAGAGGCAATGTCCTTAAGGCCAACAAGATCTCA	4877

```

? RESULT 3
? US-09-965-553-21
? Sequence 21, Application US/09965553
? Patent No. US20020112259A1
? GENERAL INFORMATION:
? APPLICANT: Wriqht, David A.
? TITLE OF INVENTION: Plant Retroelements and Methods Related Thereto
? FILE REFERENCE: P-1065 ISURF Plant Retroelement
? CURRENT APPLICATION NUMBER: US/09/965,553
? CURRENT FILING DATE: 2001-09-27
? PRIOR APPLICATION NUMBER: 09/322,478
? PRIOR FILING DATE: 1999-05-28
? PRIOR APPLICATION NUMBER: 60/087125
? PRIOR FILING DATE: 1998-05-29
? NUMBER OF SEQ ID NOS: 41
? SOFTWARE: PatentIn Ver. 2.0
? SEQ ID NO 21
? LENGTH: 4609
? TYPE: DNA
? ORGANISM: Glycine max
? US-09-965-553-21

```

	Query Match	95.2%	Score 571.2;	DB 10;	Length 4609;
	Best Local Similarity	97.0%;	Pred. No. 1.3e-186;		
	Matches 582; Conservative	0;	Mismatches 18;	Indels 0;	Gaps 0;
OY	1 TTGGAGGCTGGGGTCATATATACCCCATCTCTTGACAGCGCTTGGGTAAGCCAGTACAGTG	60			
Db	1174 TTGGAGGTTGGGCTCATATATACCCTATCTGTACAAACCTTGGGTAAACCCCGTAGCAGGTG	1233			
OY	61 GTTCCCAAGAAAGGTGGATGCACAGTGGTACGAATGAGAGGANTGACTTGATCCACA	120			
Db	1234 GTTCCCAAGAAAGGTGGATGCACAGTGGTACGAATGAGAGGANTGACTTGATCCACA	1293			
OY	121 CGAAGTGTCACTGGTTGGCGAATGTATGCATATCGCAAGCTGAATGAAGCCACACGG	180			
Db	1294 CGAAGTGTCACTGGCTGGCGAATGTATTGTACATCACAAAGCTGAATGAAGCTACACGG	1353			
OY	181 AAGGACATTTCCCCTTACCTTCATGGATGCAGTACGTTGGAGANACTTGCAGGGCAGA	240			
Db	1354 AAGGACATTTCCCCTTACCTTCATGGATGCAGTACGTTGGAGAACCTTGCAGGGCAGA	1413			
OY	241 TATACACTGTTCTTGATGATGATACTCGGGATACAAACAGATCGGGGTAGACCCAGAGAT	300			
Db	1414 TATACACTGTTCTTGATGATGATACTCGGGATACAAACAGATCGGGGTAGACCCAGATGAT	1473			

OY 301 CAGGAGAAAGCGGCTTACATGCCCCCTTGGCGTCTTTGCTTACAGAAAGATGCCATTC 360
DB 1474 CAGGAGAAAGCGGCTTACATGCCCCCTTGGCGTCTTTGCTTACAGAAAGATGCCATTC 1533
OY 361 GGGTTATGTATGACACCGACCATTTTCAGAGGTGCATGCTGCCATTTTTCAGACATG 420
DB 1534 GGGTTATGTATGACACCGACCATTTTCAGAGGTGCATGCTGCCATTTTTCAGACATG 1593
OY 421 GTGGAGAAAGCATGAGGTATTTATGACAGACTTCTCGGTTTTTGGACCCCTCATTTGAC 480
DB 1554 GTGGAGAAAGCATGAGGTATTTATGACAGACTTCTCGGTTTTTGGACCCCTCATTTGAC 1653
OY 481 AGCTGTTGAGAGACCTAGAGAGGTACTTTCAGAGGTGCAGAGACTTAATTGGTACTG 540
DB 1654 AGCTGTTGAGAGACCTAGAGAGGTACTTTCAGAGGTGCAGAGACTTAATTGGTACTG 1713
OY 541 AATTGGAGAAAGTGCATTTTCATGTTTCGAGAGGGCATAGTCTTAGGCCCAAGATCTCA 600
DB 1714 AATTGGAGAAAGTGCATTTTCATGTTTCGAGAGGGCATAGTCTTAGGCCCAAGATCTCA 1773

RESULT 4
US-09-965-553-34
Sequence 34, Application US/09965553
Patent No. US20020112259A1
GENERAL INFORMATION:
APPLICANT: Wright, David A.
APPLICANT: Voytas, Daniel F.
TITLE OF INVENTION: Plant Retroelements and Methods Related Thereto
FILE REFERENCE: P-1065 ISURF Plant Retroelement
CURRENT APPLICATION NUMBER: US/09/965,553
PRIOR FILING DATE: 2001-09-27
PRIOR APPLICATION NUMBER: 09/322,478
PRIOR FILING DATE: 1999-05-28
PRIOR APPLICATION NUMBER: 60/087125
PRIOR FILING DATE: 1998-05-29
NUMBER OF SEQ ID NOS: 41
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 34
LENGTH: 597
TYPE: DNA
ORGANISM: Glycine max
US-09-965-553-34

Query Match 94.7%; Score 568.2; DB 10; Length 597;
Best Local Similarity 97.0%; Pred. No. 4.9e-186;
Matches 579; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

OY 1 TTGGAGGCTGGGCTCATATATACCCCATCTCTGACAGCGCTTGGTAAGCCAGTACAGGTG 60
DB 1 TTGGAGGCTGGGCTCATATATACCCCATCTCTGACAGCGCTTGGTAAGCCAGTACAGGTG 60
OY 61 GTTCCCAAGAAAGGTGAATGACAGTGTACAGAGATGAGAGAAATGACTTGATACCAACA 120
DB 61 GTTCCCAAGAAAGGTGAATGACAGTGTACAGAGATGAGAGAAATGACTTGATACCAACA 120
OY 121 CGAAGTGCACCTGTTGGCGAATGTGTATCGACTATCGCAAGCTGATGAAGCCACACGG 180
DB 121 CGAAGTGCACCTGTTGGCGAATGTGTATCGACTATCGCAAGCTGATGAAGCTGATGAAGCCACACGG 180
OY 181 AAGGACATTTTCCCTTACCTTTCATGATGATCTGAGAGACTTGGACGGGACAGCA 240
DB 181 AAGGACATTTTCCCTTACCTTTCATGATGATCTGAGAGACTTGGACGGGACAGCA 240
OY 241 TACTACTGTTTCTTGATGATGATCTCGGATACAAACAGATCGCGGTAGACCCAGAGAT 300
DB 241 TACTACTGTTTCTTGATGATGATCTCGGATACAAACAGATCGCGGTAGACCCAGAGAT 300
OY 301 CAGGAGAAAGCGGCTTACATGCCCCCTTGGCGTCTTTGCTTACAGAAAGATGCCATTC 360
DB 301 CAGGAGAAAGCGGCTTACATGCCCCCTTGGCGTCTTTGCTTACAGAAAGATGCCATTC 360

OY 361 GGGTTATGTATGACACCGACCATTTTCAGAGGTGCATGCTGCCATTTTTCAGACATG 420
DB 361 GGGTTATGTATGACACCGACCATTTTCAGAGGTGCATGCTGCCATTTTTCAGACATG 420
OY 421 GTGGAGAAAGCATGAGGTATTTATGACAGACTTCTCGGTTTTTGGACCCCTCATTTGAC 480
DB 421 GTGGAGAAAGCATGAGGTATTTATGACAGACTTCTCGGTTTTTGGACCCCTCATTTGAC 480
OY 481 AGCTGTTGAGAGACCTAGAGAGGTACTTTCAGAGGTGCAGAGACTTAATTGGTACTG 540
DB 481 AGCTGTTGAGAGACCTAGAGAGGTACTTTCAGAGGTGCAGAGACTTAATTGGTACTG 540
OY 541 AATTGGAGAAAGTGCATTTTCATGTTTCGAGAGGGCATAGTCTTAGGCCCAAGATC 597
DB 541 AATTGGAGAAAGTGCATTTTCATGTTTCGAGAGGGCATAGTCTTAGGCCCAAGATC 597

RESULT 5
US-09-965-553-19
Sequence 19, Application US/09965553
Patent No. US20020112259A1
GENERAL INFORMATION:
APPLICANT: Wright, David A.
APPLICANT: Voytas, Daniel F.
TITLE OF INVENTION: Plant Retroelements and Methods Related Thereto
FILE REFERENCE: P-1065 ISURF Plant Retroelement
CURRENT APPLICATION NUMBER: US/09/965,553
PRIOR FILING DATE: 2001-09-27
PRIOR APPLICATION NUMBER: 09/322,478
PRIOR FILING DATE: 1999-05-28
PRIOR APPLICATION NUMBER: 60/087125
PRIOR FILING DATE: 1998-05-29
NUMBER OF SEQ ID NOS: 41
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 19
LENGTH: 9829
TYPE: DNA
ORGANISM: Glycine max
US-09-965-553-19

Query Match 88.7%; Score 532; DB 10; Length 9829;
Best Local Similarity 94.0%; Pred. No. 6.8e-173;
Matches 564; Conservative 0; Mismatches 35; Indels 1; Gaps 1;

OY 1 TTGGAGGCTGGGCTCATATATACCCCATCTCTGACAGCGCTTGGTAAGCCAGTACAGGTG 60
DB 4271 TTGGAGGCTGGGCTCATATATACCCCATCTCTGATAGCGCTTGGTAAGTCCAGTACAGGTG 4330
OY 61 GTTCCCAAGAAAGGTGAATGACAGTGTACAGAGATGAGAGAAATGACTTGATACCAACA 120
DB 4331 GTTCCCAAGAAAGGTGAATGACAGTGTACAGAGATGAGAGAAATGACTTGATACCAACA 4390
OY 121 CGAAGTGCACCTGTTGGCGAATGTGTATCGACTATCGCAAGCTGATGAAGCCACACGG 180
DB 4391 CGAAGTGCACCTGTTGGCGAATGTGTATCGACTATCGCAAGCTGATGAAGCCACACAG 4450
OY 181 AAGGACATTTTCCCTTACCTTTCATGATGATCTGAGAGACTTGGACGGGACAGCA 240
DB 4451 AAGGACATTTTCCCTTACCTTTCATGATGATCTGAGAGACTTGGACGGGACAGCA 4510
OY 241 TACTACTGTTTCTTGATGATGATCTCGGATACAAACAGATCGCGGTAGACCCAGAGAT 300
DB 4511 TACTACTGTTTCTTGATGATGATCTCGGATACAAACAGATCGCGGTAGACCCAGAGAT 4569
OY 301 CAGGAGAAAGCGGCTTACATGCCCCCTTGGCGTCTTTGCTTACAGAAAGATGCCATTC 360
DB 4570 CAGGAGAAAGCGGCTTACATGCCCCCTTGGCGTCTTTGCTTACAGAAAGATGCCATTC 4629
OY 361 GGGTTATGTATGACACCGACCATTTTCAGAGGTGCATGCTGCCATTTTTCAGACATG 420
DB 4630 GGGTTATGTATGACACCGACCATTTTCAGAGGTGCATGCTGCCATTTTTCAGACATG 4689
OY 421 GTGGAGAAAGCATGAGGTATTTATGACAGACTTCTCGGTTTTTGGACCCCTCATTTGAC 480

```

Db      4690 GTGAGAGAGACATGAGATATTTATGAGACGACTTCTGGATTTTGGACCCATATTGAC 4749
Oy      481 AGCTGTTAGAGAACTAGAGAGGTACTTGCAGAGGTGGAGAGACTAATCTTGAC 540
Db      4750 AACTATTGAGAACTAGAGATGCTACTACAGAGTGGTATGACTTACTTGGTACTA 4809
Oy      541 AATTGGAGAAAGTGCATTTTCATGCTTGCAGAGGAGCAATAGTCTTACGCCACAATCTCA 600
Db      4810 AATTGGAGAAAGTGCATTTTCATGCTTGCAGAGGAGCAATAGTCTTACGCCACAATCTCA 4869

```

RESULT 6

```

US-09-965-553-20
; Sequence 20, Application US/09965553
; Patent No. US20020112259A1
; GENERAL INFORMATION:
; APPLICANT: Wright, David A.
; APPLICANT: Voytas, Daniel F.
; TITLE OF INVENTION: Plant Retroelements and Methods Related Thereto
; FILE REFERENCE: P-1065 ISURF Plant Retroelement
; CURRENT APPLICATION NUMBER: US/09/965,553
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 09/322,478
; PRIOR FILING DATE: 1999-05-28
; PRIOR APPLICATION NUMBER: 60/087125
; PRIOR FILING DATE: 1998-05-29
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 20
; LENGTH: 12571
; TYPE: DNA
; ORGANISM: Glycine max
US-09-965-553-20

```

Query Match 87.2%; Score 523.2; DB 10; Length 12571;
 Best Local Similarity 93.2%; Pred. No. 8.5e-170;
 Matches 559; Conservative 0; Mismatches 38; Indels 3; Gaps 1;

```

Oy      1 TTGGAGGCTGGGCTCATATACCCATCTCTGACAGCGGTGGTAAGCCAGTACAGGTG 60
Db      4295 TTGGAGGCTGGGCTCATATACCCCTCTCTGACAGCTGCTGGTAAGCCAGTACAGGTG 4354
Oy      61 GTTCCCAAGAAAGGTGAATGACAGTGTGACAGATGAGAGAAATGACTTGTATACCA 120
Db      4355 GTTCCCAAGAAAGGTGAATGACAGTGTGACAGAAATGAGAAATGACTTGTATACCA 4414
Oy      121 CGAAGCTGCTACTGTTGGCGCAATGTATCGACTATCGCAAGCTGATGAGAGCCAG 180
Db      4415 CGAAGCTGCTACTGTTGGCGCAATGTATCAACTATCGCAAGCTGATGAGAGCCAG 4474
Oy      181 AAGGACCATTTCCCTTACCTTTCATGATCAGATGCTGGAGAGACTTGCAGAGGAG 240
Db      4475 AAGGACCATTTCCCTTACCTTTCATGATCAGATGCTGGAGAGACTTGCAGAGGAG 4534
Oy      241 TACTACTGTTCTTGGATGATGATCTCGGATACCAACAGATCGGTAGACCCAGAG 300
Db      4535 TACTACTGTTCTTGGATGATGATCTCGGATATATAATCAGATGCGGTAGACCCAG 4594
Oy      301 CAGGAGAAAGCGGCTTTTACATGCCCCCTTGGCGCTTGTGCTTACAGAAAGATGC 360
Db      4595 CAGGAGAAAGCGGCTTTTACATGCCCCCTTGGCGCTTGTGCTTACAGAAAGATGC 4654
Oy      361 GGGTTATGATGACAGGACCATTTTCAGAGGTGATGCTGGCCATTTTTCAGACATG 420
Db      4655 GGGTTATGATGACAGGACCATTTTCAGAGGTGATGCTGGCCATTTTTCAGACATG 4714
Oy      421 GTGGAGAAAGCATCGAGATATTTATGAGACGACTTCTGCTTTTGGAGCCCTCATTTG 480
Db      4715 GTGTAGAAAGCATCGAGATATTTATGAGACGACTTCTGCTTTTGGAGCCCTCATTTG 4774
Oy      481 AGCTGTTTGGAGAACTAGAGAGGCTTACAGAGGTGCGAAGAGACTTACTTGGTACTG 540

```

```

Db      4775 A---GTTTGAAGAACTAGAGATGCTACTTTAGAGTGTAGAGACTAATCTTGACTG 4831
Oy      541 AATTGGAGAAAGTGCATTTTCATGCTTGCAGAGGAGCAATAGTCTTACGCCACAATCTCA 600
Db      4832 AACTGGAGAAAGTGCATTTTCATGCTTGCAGAGGAGCAATAGTCTTACGCCACAATCTCA 4891

```

RESULT 7

```

US-09-965-553-22
; Sequence 22, Application US/09965553
; Patent No. US20020112259A1
; GENERAL INFORMATION:
; APPLICANT: Wright, David A.
; APPLICANT: Voytas, Daniel F.
; TITLE OF INVENTION: Plant Retroelements and Methods Related Thereto
; FILE REFERENCE: P-1065 ISURF Plant Retroelement
; CURRENT APPLICATION NUMBER: US/09/965,553
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 09/322,478
; PRIOR FILING DATE: 1999-05-28
; PRIOR APPLICATION NUMBER: 60/087125
; PRIOR FILING DATE: 1998-05-29
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 22
; LENGTH: 9139
; TYPE: DNA
; ORGANISM: Glycine max
US-09-965-553-22

```

Query Match 57.6%; Score 345.6; DB 10; Length 9139;
 Best Local Similarity 73.5%; Pred. No. 2e-108;
 Matches 441; Conservative 0; Mismatches 159; Indels 0; Gaps 0;

```

Oy      1 TTGGAGGCTGGGCTCATATACCCATCTCTGACAGCGGTGGTAAGCCAGTACAGGTG 60
Db      1646 TAGGAGCAGGCTTATTTACCCCTCTCGGATATGATGGTGAAGCTTGTGAGGTT 1705
Oy      61 GTTCCCAAGAAAGGTGAATGACAGTGTGACAGATGAGAGAAATGACTTGTATACCA 120
Db      1706 GTTCCCAAGAAAGGTGAATGACAGTGTGATTAATGATTAAGATGATTAATATCCAT 1765
Oy      121 CGAAGCTGCTACTGTTGGCGCAATGTATCGACTATCGCAAGCTGATGAGAGCCAG 180
Db      1766 AGGACTGCTACCGGCTGAGAGATGCTATCGAATCGAAGCTGAATGATGCCACTGG 1825
Oy      181 AAGGACCATTTCCCTTACCTTTCATGATCAGATGCTGGAGAGACTTGCAGAGGAG 240
Db      1826 AAGGACCATTTACCTTTCATGATGAGACCAAAATGCTTGAAGAGCTTGTAGGTAATCC 1885
Oy      241 TACTACTGTTCTTGGATGATGATCTCGGATACCAACAGATGCGGTAGACCCAGAGAT 300
Db      1886 TATTATTGTTTTCGATGATGATCTGCTATATTAATGATGTTGTTGATGCTTAAGAT 1945
Oy      301 CAGGAGAAAGCGGCTTTTACATGCCCCCTTGGCGCTTGTGCTTACAGAAAGATGC 360
Db      1946 CAGGAGAAAGCGGCTTTTACATGCCCCCTTGGCGCTTGTGATATCGGACATGCTTTT 2005
Oy      361 GGGTTATGATGACAGGACCATTTTCAGAGGTGATGCTGGCCATTTTTCAGACATG 420
Db      2006 GGGTTATGATGACAGGACCATTTTCAGAGGTGATGCTGGCCATTTTTCAGATG 2065
Oy      421 GTGGAGAAAGCATCGAGATATTTATGAGACGACTTCTGCTTTTGGAGCCCTCATTTG 480
Db      2066 GTGGAGAAAGCATCGAGATATTTATGAGATATTTCTATTTTGGGCACTCTTTAAG 2125
Oy      481 AGCTGTTTGAAGAACTAGAGAGGCTTACAGAGGTGCGAAGAGACTTACTTGGTACTG 540
Db      2126 GGGGCTTATTAATCTTGAAGAGATTTACAGAGATGTAAGATGCAATCTAGTCTC 2185
Oy      541 AATTGGAGAAAGTGCATTTTCATGCTTGCAGAGGCTTACAGAGGCTTACAGAGGCTTAC 600
Db      2186 AATTGGAGAAAGTGCATTTTCATGCTTGCAGAGGCTTACAGAGGCTTACAGAGGCTTAC 2245

```

RESULT 8

US-09-965-553-29

; Sequence 29, Application US/09965553

; Patent No. US20020112259A1

; GENERAL INFORMATION:

; APPLICANT: Wright, David A.

; APPLICANT: Voytas, Daniel F.

; TITLE OF INVENTION: Plant Retroelements and Methods Related Thereto

; FILE REFERENCE: P-1065 ISURF Plant Retroelement

; CURRENT APPLICATION NUMBER: US/09/965,553

; CURRENT FILING DATE: 2001-09-27

; PRIOR APPLICATION NUMBER: 09/322,478

; PRIOR FILING DATE: 1999-05-28

; PRIOR APPLICATION NUMBER: 60/087125

; PRIOR FILING DATE: 1998-05-29

; NUMBER OF SEQ ID NOS: 41

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 29

; LENGTH: 597

; TYPE: DNA

; ORGANISM: Pisum sativum

US-09-965-553-29

Query Match 52.3%; Score 313.8; DB 10; Length 597;

Best Local Similarity 70.4%; Pred. No. 5e-98;

Matches 420; Conservative 0; Mismatches 177; Indels 0; Gaps 0;

QY 1 TTGAGGCTGGGCTCATATATACCCCATCTGTACAGCGGCTGGGTAAGCCAGTACAGTG 60
 DB 1 TTGAGTGGAGATGATTTACCCGATCTGGATGATGATGATGATGATGATGATGATGATG 60
 QY 61 GTTCCCAAGAAAGGTGGAAATGACAGTGTGACAGATGAGAGAAATGATGATGATGATG 120
 DB 61 GTTCCCAAGAAAGGTGGAAATGACAGTGTGACAGATGAGAGAAATGATGATGATGATG 120
 QY 121 CGAATGTACATGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGG 180
 DB 121 CGAATGTACATGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGG 180
 QY 181 AAGGACATTTCCCTTACCTTTCATGATGATGATGATGATGATGATGATGATGATGATG 240
 DB 181 AAGGACATTTCCCTTACCTTTCATGATGATGATGATGATGATGATGATGATGATGATG 240
 QY 241 TACTACTGTTTCTTGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
 DB 241 TACTACTGTTTCTTGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
 QY 301 CAGGAGAGAGCGGCTTTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
 DB 301 CAGGAGAGAGCGGCTTTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
 QY 361 GGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 420
 DB 361 GGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 420
 QY 421 GTGGAGAAAGAGTGAAGTATTTATGAGAGACTTTCGCTTTTGGACCTTAC 480
 DB 421 GTGGAGAAAGAGTGAAGTATTTATGAGAGACTTTCGCTTTTGGACCTTAC 480
 QY 481 AGCTTTTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGA 540
 DB 481 AGCTTTTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGA 540
 QY 541 AATTTGAGAAAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGA 597
 DB 541 AATTTGAGAAAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGA 597

RESULT 9

US-09-965-553-27

; Sequence 27, Application US/09965553

; Patent No. US20020112259A1

; GENERAL INFORMATION:

; APPLICANT: Wright, David A.

; APPLICANT: Voytas, Daniel F.

; TITLE OF INVENTION: Plant Retroelements and Methods Related Thereto

; FILE REFERENCE: P-1065 ISURF Plant Retroelement

; CURRENT APPLICATION NUMBER: US/09/965,553

; CURRENT FILING DATE: 2001-09-27

; PRIOR APPLICATION NUMBER: 09/322,478

; PRIOR FILING DATE: 1999-05-28

; PRIOR APPLICATION NUMBER: 60/087125

; PRIOR FILING DATE: 1998-05-29

; NUMBER OF SEQ ID NOS: 41

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 27

; LENGTH: 600

; TYPE: DNA

; ORGANISM: Arabidopsis thaliana

US-09-965-553-27

Query Match 50.8%; Score 304.6; DB 10; Length 600;

Best Local Similarity 69.3%; Pred. No. 7.7e-95;

Matches 415; Conservative 0; Mismatches 184; Indels 0; Gaps 0;

QY 2 TGGAGGCTGGGCTCATATATACCCCATCTGTACAGCGGCTGGGTAAGCCAGTACAGTG 61
 DB 2 TGGAGTGGAGATGATTTACCCGATCTGGATGATGATGATGATGATGATGATGATGATG 61
 QY 62 TTCCCAAGAAAGGTGGAAATGACAGTGTGACAGATGAGAGAAATGATGATGATGATG 121
 DB 62 TTCCCAAGAAAGGTGGAAATGACAGTGTGACAGATGAGAGAAATGATGATGATGATG 121
 QY 122 GAATGTCTGCTGGTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGG 181
 DB 122 GAATGTCTGCTGGTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGG 181
 QY 182 AGGACATTTCCCTTACCTTTCATGATGATGATGATGATGATGATGATGATGATGATG 241
 DB 182 AGGACATTTCCCTTACCTTTCATGATGATGATGATGATGATGATGATGATGATGATG 241
 QY 242 ACTACTGTTTCTTGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 301
 DB 242 ACTACTGTTTCTTGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 301
 QY 302 AAGGAGAGAGCGGCTTTCATGATGATGATGATGATGATGATGATGATGATGATGATG 361
 DB 302 AAGGAGAGAGCGGCTTTCATGATGATGATGATGATGATGATGATGATGATGATGATG 361
 QY 362 GGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 421
 DB 362 GGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 421
 QY 422 TGGAGAAAGAGTGAAGTATTTATGAGAGACTTTCGCTTTTGGACCTTAC 481
 DB 422 TGGAGAAAGAGTGAAGTATTTATGAGAGACTTTCGCTTTTGGACCTTAC 481
 QY 482 GCTGTTTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGA 541
 DB 482 GCTGTTTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGA 541
 QY 542 ATTGGAGAAAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGA 600
 DB 542 ATTGGAGAAAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGA 600

RESULT 10

US-09-965-553-23

; Sequence 23, Application US/09965553

; Patent No. US20020112259A1

; GENERAL INFORMATION:

; APPLICANT: Wright, David A.

; APPLICANT: Voytas, Daniel F.

; TITLE OF INVENTION: Plant Retroelements and Methods Related Thereto

D5 405 CCATCGCAGCAATTCTATATATGATGACTTAGTAGTAATAGGTTGTTCCAGAAAAACAT 464

QY	DB	QY	DB
484	484	544	544
TGTTTGAGGAACCTFAGAGAGGGGTACTCTTCAGAGTGGGAGAGACGACTTAATTGGTACGAT	484	TGGGAAAAGTGCATTTCTTATCGTGTGAGAGGGGCATATGCTCTAGGCCACAA	544
543	543	593	593
465	465	525	525
ATGCTCAAAAATTTTACCTAACCTATTTCCGACGATATGAGACGACATATTTGAAACCTACAT	465	CCAGGGAATATGTCCTTTCTTATGAAAAGAACTATATTTGGGTACCA	525
524	524	574	574

```

RESULT 13
US-09-997-672-1
Sequence 1, Application US/09997672
Publication No. US20030061632A1
GENERAL INFORMATION:
APPLICANT: Weterings, Koen
APPLICANT: Apuya, Nestor R.
APPLICANT: Tatarkinova, Tatiana
APPLICANT: Goldberg, Robert B.
APPLICANT: The Regents of the University of California
APPLICANT: Ceres, Inc.
TITLE OF INVENTION: Polynucleotides Useful for Modulating Transcription
FILE REFERENCE: 023070-115810US
CURRENT APPLICATION NUMBER: US/09/997,672
CURRENT FILING DATE: 2001-11-28
PRIOR APPLICATION NUMBER: US 60/253,672
PRIOR FILING DATE: 2000-11-28
NUMBER OF SEQ ID NOS: 42
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 4298
TYPE: DNA
ORGANISM: Phaseolus coccineus
FEATURE:
NAME/KEY: promoter
LOCATION: (1)..(4298)
OTHER INFORMATION: Scarlet Runner Bean G654 promoter
NAME/KEY: modified base
LOCATION: (1)..(4298)
OTHER INFORMATION: n = g, a, c or t
US-09-997-672-1

```

	Query Match	Similarity	10.7%	Score 64	DB 9	Length 4282
	Best Local	Similarity	46.3%	Pred. No. 3.6e-11		
	Matches	211	Conservative	0	Mismatches 245	Indels 0
QY	105	TGACCTGATACCAACAGCACTGTCTACGTGTGGGCAATGTATACGACTATCGCAACT	164			
Db	276	TGCTTTTGTGGTGCACAAAAAAGATGAAAAATGGCGTATGTGTGTATTTGAGAGCAAT	335			
QY	165	GATGAAAGCCACAGGAAAGACATTTCCCTTACCTTTCATGTATCGATGCTGGAAG	224			
Db	336	CACACATCTCCATCCAGTATAGGCAATCCATCCAAAGGCTTGACGATATGCTTGAATGA	395			
QY	225	ACTTGACGGGACAGGCATCTACTGTCTTTCTTGAGATGATCTCGGATATCAACACAGATCGC	284			
Db	396	ATTCGATGGGTCACACTCTATCTCCAAATTTGACCTTAAAGTATATACCAAAATTCG	455			
QY	285	GGTGAACCCACAGATCAGAGAAAGCGCCTTTACATGCCCTTTGCGCTTGTGCTTA	344			
Db	456	AATCAAGAGAGGTATGATGATGAAACCGCTTTTAAAGCAAAATTTGATTATATAGAGTG	515			
QY	345	CAGAAGATGCGATTCGGGTATGTAAATGACACCGCCACATTTGAGAGGTGCATCTCGC	404			
Db	516	GTGTGTATGTCCTTTGGTCTTCTTACTTAAACGCTCCAGATCATTCATGAGGCTTATGAAATCA	575			
QY	405	CATTTTTTCACACATGTGTGAGAAAGACATCGAGATTTTATGACGACTCTTCGGTTT	464			
Db	576	CACCTTGAGGATTTGTATAGTAAATATGTAGTATTTTGTATGATATCTTAGTATA	635			
QY	465	TGGAACCTCATTTTACAGCTGTTTGAGAGAACCTTAGAAGGCTACTTCAGAGGTGGAAGA	524			
Db	636	TATGTAAGAACCTTAAGACATCTTAAGTACACCTTAAAGGAAGCTTCTTCAAGTTCTTAAAGAA	695			

```

OY      525 GACTAAGCTTGGTACTGAAATGGGAAAGTGTCATT 560
          | | | | | | | | | |
DB      696 AATAGTCTTTTGGCAATAGGATAGTGACCTT 731

```

```

US-09-997-672-2
; Sequence 2, Application US/09997672
; Publication No. US20030061632A1
; GENERAL INFORMATION:
; APPLICANT: Weterings, Koen
; APPLICANT: Apuya, Nestor R.
; APPLICANT: Tatatinova, Tatiana
; APPLICANT: Goldberg, Robert B.
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Polynucleotides Useful for Modulating Transcription
; FILE REFERENCE: 023070-115810US
; CURRENT APPLICATION NUMBER: US/09/997,672
; CURRENT FILING DATE: 2001-11-28
; PRIOR APPLICATION NUMBER: US 60/253,672
; PRIOR FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4921
; TYPE: DNA
; ORGANISM: Phaseolus coccineus
; FEATURE:
; OTHER INFORMATION: Scarlet Runner Bean G654 genomic region
; NAME/KEY: exon
; LOCATION: (4299)..(4346)
; NAME/KEY: Intron
; LOCATION: (4347)..(4509)
; NAME/KEY: exon
; LOCATION: (4510)..(4734)
; NAME/KEY: modified_base
; LOCATION: (1)..(4921)
; OTHER INFORMATION: n = g, a, c o r t
US-09-997-672-2

```

Query Match	10.7%	Score 64;	DB 9;	Length 4921;
Best Local Similarity	46.3%;	Pred. No. 3.9e-11;		
Matches 211;	Conservative	0;	Mismatches 245;	Indels 0;
			Gaps	0
QY	105 TGACTTGATACACACGACGACACTGTCTACTGGTTGGCGGAATGTGTATCGACTATCGCAAGCT	164		
DB	276 TGTCTTTGTGGTCCAAAAAAGATGGAATAAGGGGTATGTGTGTGATTTGATGAGCAAT	335		
QY	165 GATTAAGCCACACGGAAGGACCAATTTCCCTTACCTTTCATNGATCAGATCCTGGAGAG	224		
DB	336 CAACCAACATCACCATCAAGTATAGCATATCCCAAGGCTTGACGATATCTTGGATGA	395		
QY	225 ACTTCAGAGCGCGGCACTACTACTGTCTTTCGTGGATGATATCTGGGATACACAGATCGC	284		
DB	366 ATGTGATGGGTCACTCTATCTTCCAAATTATACCTTAAAGTGGATATACCACAATTGC	455		
QY	285 GGTAGACCCACAGATCAGAGAGACGCGCTTTACATGCCCCCTTGGCGTTCCTTA	344		
DB	456 AATCAAGAGGGGTGATGATGGAATAACCGCTTTTAAAGACCAATTGTGATTTATGAGTG	515		
QY	345 CAGAAGGATCGCATTCGGGGTATATGTAATGCAACCACCACCAATTTGAGAGGTGATCCTGGC	404		
DB	516 GTTGGTGAAGCCCTTTGGTCTTACTACACGCTCCAAAGTACATTTCATGAGGCTTATAAATCA	575		
QY	405 CATTTTTTCACACATGTGAGAGAAAGACATGAGAGGTATTTATGACGACTTCGCTTTT	464		
DB	576 CACCTTGAGGGATTTGTATAGTAATAATATGTAGTATTTATTTTGTATGATATCTTATGATA	635		
QY	465 TGGACCTTCATTTTGACAGCTGTTTGAGAGACCTTAGAGAGGGTACTTCAGAGGTCGGAAGA	524		
DB	636 TAGTAAACCCCTAGAGACCATCTAATGACACCTTTGGGAGAGTCTTCTAGTCTTAGGAA	695		

OY 525 GACTACTGCTACTGAATGGGAAAGTGCATT 560
| | | | | | | | | | | | | | | | | |
Db 696 AAATAGTCTTTTGGCAATAGGAGATAGTACTT 731

RESULT 15

US-09-864-680-1/c
; Sequence 1, Application US/09864680
; Patent NO. US20020012981A1
; GENERAL INFORMATION:
; APPLICANT: Staskiewicz, et al.
; TITLE OF INVENTION: Bst Resistance Gene
; FILE REFERENCE: 50687
; CURRENT APPLICATION NUMBER: US/09/864,680
; CURRENT FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 09/360,186
; PRIOR FILING DATE: 1999-07-23
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 31491
; TYPE: DNA
; ORGANISM: Capsicum annuum
US-09-864-680-1

Query Match 8.1%; Score 48.6; DB 10; Length 31491;

Best Local Similarity 46.8%; Pred. No. 2.1e-05;

Matches 153; Conservative 0; Mismatches 174; Indels 0; Gaps 0;

OY 139 CGAATGTGTCGACCTTCGCAAGCTGATGATGAGCCACAGGAGACCATTCCCTTA 198
| | | | | | | | | | | | | | | | | |
Db 24307 CGGATGTGCATAGACTCCGTCAGTTGAAATGATGATGATTAATAATATATCCTT 24248
| | | | | | | | | | | | | | | | | |
OY 199 CCTTCATGGATGATGCTGAGAGACTTCAGGGGAGCATACTACTGTTCTTGAT 258
| | | | | | | | | | | | | | | | | |
Db 24247 CCTAGATTTGATGACCTTTTTCAGCAGCTTCAGGGTCTAAGTCTTTCAAAAAATAGAC 24188
| | | | | | | | | | | | | | | | | |
OY 259 GGATGCTGGGATGACAGATGCGGTAGACCCAGATCAGAGAGAGAGGCGCTT 318
| | | | | | | | | | | | | | | | | |
Db 24187 CTTCGTCGGGTTCATGACATGATGAAATAGGGAGTCAACATACCAAGACAGCCTTC 24128
| | | | | | | | | | | | | | | | | |
OY 319 ACATGCCCCCTTGGCGCTTGTCTACAGAGATGCCATTGGGTTATGTAATGCACCA 378
| | | | | | | | | | | | | | | | | |
Db 24127 CGAACCAGATGCTCACTCAAAATTTTATGTCATGTCCTTCGGGTGATTAACGCCCT 24068
| | | | | | | | | | | | | | | | | |
OY 379 GCCAATTTTCAGAGbTGCATGCTGCCATTTTTCAGACATGATGAGAGAAAAGCATCGAG 438
| | | | | | | | | | | | | | | | | |
Db 24067 GCAGCCTTTCATGATCTTATGAGTAGAGTGTCCGTCAATTATGACTTGTTCGTCAAT 24008
| | | | | | | | | | | | | | | | | |
OY 439 GATTTATGAGAGACTTCTCGGTTTTT 465
| | | | | | | | | | | | | | | | | |
Db 24007 GATTTATGATATATATTCATCTAT 23981
| | | | | | | | | | | | | | | | | |

Search completed: June 21, 2003, 00:34:41
Job time : 222.698 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2003, 18:49:21 ; Search time 2303.77 Seconds
(without alignments)
4217.989 Million cell updates/sec

Title: US-09-965-553-11
Perfect score: 600
Sequence: 1 ttggaagctggcctcatata.....tccatagccacaagatctca 600

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

- Database :
- 1: em_estba:*
 - 2: em_esthum:*
 - 3: em_estin:*
 - 4: em_estnu:*
 - 5: em_estov:*
 - 6: em_estpl:*
 - 7: em_estro:*
 - 8: em_hlc:*
 - 9: gb_estl:*
 - 10: gb_est2:*
 - 11: gb_hlc:*
 - 12: gb_est3:*
 - 13: gb_est4:*
 - 14: gb_est5:*
 - 15: em_estfun:*
 - 16: em_estom:*
 - 17: gb_gss:*
 - 18: em_gss_hum:*
 - 19: em_gss_inv:*
 - 20: em_gss_pln:*
 - 21: em_gss_vit:*
 - 22: em_gss_fun:*
 - 23: em_gss_mam:*
 - 24: em_gss_mus:*
 - 25: em_gss_other:*
 - 26: em_gss_pro:*
 - 27: em_gss_rnd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	317	52.8	582	17	BH421575 BOHEX22TF
2	315.2	52.5	727	17	BH423185 BOHHA57TR
3	315.2	52.5	841	17	BH718174 BOHGI87TR
4	310.4	51.7	759	17	BH655348 BOMEJ58TR
5	310.4	51.7	834	17	BH435343 BOHAU19TR
6	308.8	51.5	762	17	BH497441 BOHBD63TR

C.	7	307.6	51.3	842	17	BH448511 BOGOM52TR
	8	307.2	51.2	768	17	BH714463 BOMPO54TR
	9	307.2	51.2	814	17	BH697068 BOMOT46TR
	10	304.6	50.8	805	17	BH597975 BOHGI95TR
	11	304.4	50.7	854	17	BH711455 BOHRT83TR
	12	304.4	50.4	743	17	BH703456 BOMGU48TR
	13	302.4	50.4	796	17	BH501651 BOGOM29TR
	14	302.4	50.3	806	17	BH443545 BOGOG29TR
	15	301.6	50.3	697	14	BO996483 OGG12P10.
	16	300.4	50.1	859	17	BH420759 BOG1223TR
	17	297.2	49.5	821	17	BH656439 BOMTR94TR
	18	296.8	49.5	802	17	BH434092 BOGZS83TR
	19	296.6	49.4	576	17	BH530743 BOHOG52TR
	20	294.4	49.0	759	17	BH682723 BOMER67TR
	21	294	49.0	766	17	BH550351 BOHAK33TR
	22	293.4	48.9	806	17	BH244881 AUIIA12TR
	23	292.6	48.8	843	17	BH448694 BOG215TR
	24	290.8	48.5	799	17	BH701377 BOMHH20TR
	25	290.8	48.5	843	17	BH54612 BOGUM67TR
	26	290.2	48.4	661	17	BH505270 BOHSE73TR
	27	289.8	48.3	854	17	BH452427 BOGOG24TR
	28	288.8	48.1	734	17	BH576493 BOHMB67TR
	29	288.4	48.1	667	17	AQ956471 LERAK19TR
	30	287.8	48.0	757	17	BH654786 BOMDG77TR
	31	287.6	47.9	866	17	BH576412 BOHQR21TR
	32	286.8	47.8	845	17	BH580975 BOGDR46TR
	33	286	47.7	681	17	BH508673 BOGPP31TR
	34	286	47.7	840	17	BH536770 BOHRN81TR
	35	285.2	47.5	444	17	BH405540 Gm_15B001
	36	284.4	47.4	744	17	BH661378 BOMNH43TR
	37	284.4	47.4	811	17	BH588426 BOHAA22TR
	38	282.8	47.1	800	17	BH543949 BOGUR87TR
	39	282	47.0	757	17	BH734399 BOMIH73TR
	40	280.4	46.7	825	17	BH479539 BOGVI70TR
	41	279.4	46.6	741	17	BH694557 BOHVI20TR
	42	279.4	46.6	760	17	BH484943 BOGLI32TR
	43	278.8	46.5	741	17	BH589626 BOGPC86TR
	44	278.4	46.4	737	17	BH709535 BOMAG18TR
	45	278.2	46.4	607	17	BH523876 BOHMT89TR

ALIGNMENTS

RESULT 1
BH421575/c 582 bp DNA linear GSS 12-DEC-2001
LOCUS BOHEX22TF BOHE Brassica oleracea genomic clone BOHEX22, DNA
DEFINITION
ACCESSION BH421575
VERSION BH421575.1 GI:17607303
KEYWORDS GSS.
SOURCE Brassica oleracea.
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 582)
TOWN,C.D., Van Aken,S., Uterback,T. and Fraser,C.M.
Whole genome shotgun sequencing of Brassica oleracea
Unpublished (2001)
JOURNML Other GSSs: BOHEX22TR
COMMENT Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: ctown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TF
Class: Sheared ends.
FEATURES
source 1..582
Location/Qualifiers

KEYWORDS	GSS.
SOURCE	Brassica oleracea.
ORGANISM	Brassica oleracea. Eunaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica. 1 (bases 1 to 841)
REFERENCE	Town,C.D., Van Aken,S.. Utterback,T. and Fraser,C.M. Whole genome shotgun sequencing of Brassica oleracea Unpublished (2001) Other_GSSs: BOMG187nr
AUTHORS	Contact: Chris Town
JOURNAL	TIJR
COMMENT	9712 Medical Center Drive, Rockville, MD 20850, USA. Tel: 301-838-3523 Fax: 301-838-0208 Email: cdtown@tigr.org DNA is from a doubled haploid provided by Tom Osborn. Seq primer: TF Class: sheared ends.
FEATURES	Location/Vnallifiers
Source	1..841 /organism="Brassica oleracea" /strain="T01000DH3" /db_xref="taxon:3712" /clone="BOMG187" /clone_1lb="Bo_2_3_kb" /note="Vector: pHD1; site_1: BstXI; 2-3 kb sheared genomic DNA inserted into pHD1 using BstXI linkers"
BASE COUNT	259 a 189 c 150 g 243 t
ORIGIN	
Query Match	52.5%; Score 315.2; DB 17; Length 841;
Best Local Similarity	70.3%; Pred. No. 4.4e-87;
Matches 422; Conservative	0; Mismatches 178; Indels 0; Gaps 0;
OY	1 TTGGAGGTGGGGCATATPACCACCTCCTGACAGCGCTTGCGTAGGCCAGTCACAGTGT 60
DB	TTAATGTGGTGTATTATCTACCCCTATCCCAATTTCTAAATGGATATCTGTGCATGTT 688
OY	61 GTTCCCAAGAAGGTGCAATGACAGTGGTAGAGATGAGAGAAATGACTGTATACCAACA 120
DB	GTGCCAAAAGGGGTATTAATAGTGAATTAATAATGCAAGATGATGATACCAACA 628
OY	121 CGAAGTGTACTGTTGGCGAATGTGATGCATCTATGCGAAGCTGAATGAAGCCACAGG 180
DB	AGAACAGTATTATGGGCATAGGATGTGTATGATTACCGCAAACGTAACTGTATCTAGA 568
OY	181 AAGNACATTTCCCCTTACCTTTCATGATGATGATAGATGCTGGAGAGACTTCGAGGACGA 240
DB	AAGATCATTTTCCACTTCATTTATATGATGATGATGCTGGAGAGACTTCGAATATATCA 508
OY	241 TACTACTGTTTTCTTGATGATGATACGCGGATACAACAGATCGCGATAGCCCCAGAGAT 300
DB	TTCTATTTCTTTCTTGATGATGATATTCAGAGATTTCTCCAGATCCCATATACATCCAATGAT 448
OY	301 CAGAGAGACGCGCTTTACATGCCCTTTGGCGTCTTTGCTTTACAGAGAGATGCCATTTC 360
DB	CAGAGAGAAACAAATTCACATTCACATTCCTATGATGATGATGATGATGATGATGATGAT 388
OY	361 GGATATCTATATGACACCGCACCATTTTATAGAGGTGATAGCTGCGCATTTTTCAGACATG 420
DB	GGTCTATCTATATGCTTCAGACACCTTTTAAAGGTGCAATGATGTCATCTTTTCTCATATCG 328
OY	421 GTGGAGAAACATATCGAGGTATTTATGAGACGCTCTCGGTTTTTGGAGCCCATTTTGAC 480
DB	ATTGAGAGATGTTTGAGAGGTGTATATGATGATGATTTTCTCCGTATGATGATCTTCGTTTCT 268
OY	481 AGCTGTTTGGAGAACCTTAGAGAGGTACTTCAAGAGTGCAGAGACTTACTTGGTACTG 540
DB	GCTTGTGTTGTCAAATTTGTGACAGGCTCTAGAGAGATGTGAAGACACCAACCTTGTGCTG 208
OY	541 AATTGGAAAAAGTGTATTTATGTTGTGAGAGGGCATATGTTCTTAGCCCAACAATCTTCA 600

[illegible]

QY 421 GTGAGAAAGACATGATTTATGACGACTTCTCGATTTTGGACCCATTTGAC 480
 DB 205 ATTGAGGACATATGAGAGTTTATGACGATTTCTGACTGAGGTTCTTATTTAGC 146
 QY 481 AGCTGTTGAGAACCTTAGAGAGGTACTTTCAGAGGTGGAGAGACTAATTGTTACTG 540
 DB 145 GAGTGGCTTGTCTAATCTGTGCAGAAAGTGTGAGAGAAAGAACTTGGTGTGA 86
 QY 541 AATTGGGAAAGATGCTTATGTTGAGAGGCGATGCTTACGACCAAGATCTCA 600
 DB 85 AATTGGGAAAGATGCTTATGTTGAGAGGCGATGCTTACGACCAAGATCTCA 26

RESULT 5
 BH435343/c
 LOCUS BOHAU19TF BOHA Brassica oleracea genomic clone BOHAU19, DNA
 DEFINITION
 ACCESSION BH435343
 VERSION BH435343.1 GI:17621064
 KEYWORDS GSS.
 SOURCE Brassica oleracea.
 ORGANISM Brassica oleracea.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
 1 (bases 1 to 834)
 Town, C.D., Van Aken, S., Utterback, T. and Fraser, C.M.
 Whole genome shotgun sequencing of Brassica oleracea
 Unpublished (2001)
 Other_GSSs: BOHAU19TR
 CONTACT: Chris Town
 TIGR

REFERENCE
 AUTHORS 9712 Medical Center Drive, Rockville, MD 20850, USA.
 TITLE Tel: 301-838-3523
 JOURNAL Fax: 301-838-0208
 COMMENT Email: cdtown@tigr.org
 DNA is from a doubled haploid provided by Tom Osborn.
 Seq primer: TF
 Class: sheared ends.
 Location/Qualifiers
 1..834
 /organism="Brassica oleracea"
 /strain="TO1000DH3"
 /db_xref="taxon:3712"
 /clone="BOHAU19"
 /clone_1fb="BOHA"
 /note="Vector: PHOS1; Site_1: BstXI; 2-3 kb sheared
 genomic DNA inserted into PHOS1 using BstXI linkers"

BASE COUNT 219 a 204 c 179 g 232 t
 ORIGIN

Query Match 51.7%; Score 310.4; DB 17; Length 834;
 Best Local Similarity 69.8%; Pred. No. 1.4e-85;
 Matches 419; Conservative 0; Mismatches 181; Indels 0; Gaps 0;

QY 1 TTGGAGGCTGGGCTCATATACCCCATCTGCAGAGCGCTGGTAAAGCCAGTACAGGTG 60
 DB 687 TTGAGTGCAGAGGCTGATCTACCAATTTTCAGAGTACTGGTGGAGCCCGTTTCATGTG 628
 QY 61 GTTCCCAAGAAAGTGGATGACAGTGTACAGATGATAGAGATGACTTATACCAACA 120
 DB 627 GTTCCCAAGAAAGTGGATGACAGTGTACAGATGATAGAGATGACTTATACCAACA 568
 QY 121 CGAAGTGTGAGTGGGATGTATGATGATGATGATGATGATGATGATGATGATGATG 180
 DB 567 AGAACAATATCAGAGGATAGATGATGATGATGATGATGATGATGATGATGATGATG 508
 QY 181 AAGGACATTTTCCCTTACCTTATGATGATGATGATGATGATGATGATGATGATGATG 240
 DB 507 AAGGACATTTTCCCTTACCTTATGATGATGATGATGATGATGATGATGATGATGATG 448

QY 241 TACTACTGTTCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
 DB 447 TACTACTGTTCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 388
 QY 301 CAGAGAGAGAGGCGCTTACATGCCCTTGGCGTCTTGGTTCAGAGAGATGCGCATTC 360
 DB 387 CAAGAGAGAGAGATTCACCTGTCGCATACAGTACTTTTGGCTTACAGAGAGATGCCCTTC 328
 QY 361 GGGTATGATATGACACACACCATTTTCAGAGGTGATGATGATGATGATGATGATGATG 420
 DB 327 GGATTTGACATGCTTCTGCGCATTTCCAGATGATGATGATGATGATGATGATGATGAT 268
 QY 421 GTGAGAAAGACATGAGAGGTATTTATGAGACATCTTCGCTTTTGGACCTCATTTAC 480
 DB 267 ATTGAGGACATTTATGAGAGGTATTTATGAGACATCTTCGCTTTTGGACCTCATTTAC 208
 QY 481 AGCTGTTGAGAACCTTAGAGAGGTACTTTCAGAGGTGGAGAGACTAATTGTTACTG 540
 DB 207 GAGTGGCTTGTCTAATCTGTGCAGAAAGTGTGAGAGAAAGAACTTGGTGTGA 148
 QY 541 AATTGGGAAAGATGCTTATGTTGAGAGGCGATGCTTACGACCAAGATCTCA 600
 DB 147 AATTGGGAAAGATGCTTATGTTGAGAGGCGATGCTTACGACCAAGATCTCA 88

RESULT 6
 BH497441
 LOCUS BOHB63TR BOHB Brassica oleracea genomic clone BOHB63, DNA
 DEFINITION
 ACCESSION BH497441
 VERSION BH497441.1 GI:17705545
 KEYWORDS GSS.
 SOURCE Brassica oleracea.
 ORGANISM Brassica oleracea.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
 1 (bases 1 to 762)
 Town, C.D., Van Aken, S., Utterback, T. and Fraser, C.M.
 Whole genome shotgun sequencing of Brassica oleracea
 Unpublished (2001)
 Other_GSSs: BOHB63TR
 CONTACT: Chris Town
 TIGR

REFERENCE
 AUTHORS 9712 Medical Center Drive, Rockville, MD 20850, USA.
 TITLE Tel: 301-838-3523
 JOURNAL Fax: 301-838-0208
 COMMENT Email: cdtown@tigr.org
 DNA is from a doubled haploid provided by Tom Osborn.
 Seq primer: TR
 Class: sheared ends.
 Location/Qualifiers
 1..762
 /organism="Brassica oleracea"
 /strain="TO1000DH3"
 /db_xref="taxon:3712"
 /clone="BOHB63"
 /clone_1fb="BOHB"
 /note="Vector: PHOS1; Site_1: BstXI; 2-3 kb sheared
 genomic DNA inserted into PHOS1 using BstXI linkers"

BASE COUNT 244 a 126 c 162 g 230 t
 ORIGIN

Query Match 51.5%; Score 308.8; DB 17; Length 762;
 Best Local Similarity 69.7%; Pred. No. 4.1e-85;
 Matches 418; Conservative 0; Mismatches 182; Indels 0; Gaps 0;

QY 1 TTGGAGGCTGGGCTCATATACCCCATCTGCAGAGCGCTGGTAAAGCCAGTACAGGTG 60
 DB 119 TTGATGACAGAGATATATATCTTATGATGATGATGATGATGATGATGATGATGATG 178
 QY 61 GTTCCCAAGAAAGTGGATGACAGTGTACAGATGATGATGATGATGATGATGATGATGAT 120

```

Db      179  GTCCCAAGAGAGGAGATTACTGTGTCACAAAATGATAGATGATACATATACACACA 238
Qy      121  CGAACTGCATCGTGTGGCGAATGTGTATGACTATCGAAGTGTGATGACGACACAGG 180
Db      239  AGAACAAATAGAGACATAGAGATGTGCATATGATTAATCGAAATTAACATCATCTAGA 238
Qy      181  AAGGACATTTCCCTTACCTTATGATGATGATGATGATGATGATGATGATGATGATGAT 240
Db      299  AAAGATCTTTTCCATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 338
Qy      241  TACTACTGTTTCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
Db      359  TATATATTTTCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 418
Qy      301  CAGAGAGAGAGGCGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
Db      419  CAAGAGAAACGACATTTACATGATGATGATGATGATGATGATGATGATGATGATGAT 478
Qy      361  GGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
Db      479  GGGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 538
Qy      421  GTGAGAAAGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480
Db      539  ATAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 598
Qy      481  AGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 540
Db      599  GCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 658
Qy      541  AATGGGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 600
Db      659  AATGGGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 718

```

```

RESULT 7
BH448511/c
LOCUS      BH448511      842 bp      DNA      linear      GSS 12-DEC-2001
DEFINITION      BOCO52TF BOCO Brassica oleracea genomic clone BOCO52, DNA
sequence.
ACCESSION      BH448511
VERSION        BH448511.1 GI:17634222
KEYWORDS      GSS.
SOURCE        Brassica oleracea.
ORGANISM      Brassica oleracea.
REFERENCE      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
AUTHORS      Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
TITLE      Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
JOURNAL      1 (bases 1 to 842)
COMMENT      Town, C.D., Van Aken, S., Uteirback, T. and Fraser, C.M.
              Whole genome shotgun sequencing of Brassica oleracea
              Unpublished (2001)
              Other_GSSs: BOCO52TF
              Contact: Chris Town
              TIGR
              9712 Medical Center Drive, Rockville, MD 20850, USA.
              Tel: 301-838-3523
              Fax: 301-838-0208
              Email: cdtown@tigr.org
              DNA is from a doubled haploid provided by Tom Osborn.
              Seq primer: TF
              Class: sheared ends.

```

```

FEATURES
SOURCE
1. 842
Location/Qualifiers
/organism="Brassica oleracea"
/strain="TO1000DH3"
/db_xref="taxon:3712"
/clone="BOCO52"
/clone_1tb="BOGO"
/note="Vector: PHOS1, Site_1: BstXI; 2-3 kb sheared
genomic DNA inserted into PHOS1 using BstXI linkers"
BASE COUNT      227 a      201 c      170 g      244 t

```

```

ORIGIN
Query Match      51.3%; Score 307.6; DB 17; Length 842;
Best Local Similarity 69.9%; Pred. No. 1e-84;
Matches 415; Conservative 0; Mismatches 179; Indels 0; Gaps 0;
Qy      7      GCTGGGCTCATATATACCCCATCTCTGTGACAGCGGTGGGTAGGCCCATGAGGTGTC 66
Db      626  GCAGGGGTGATCTACCCCATTTTCAAGACAGCATTTGGGTAGGCCGTTCTATGTGTTCT 567
Qy      67      AAGAAAGGTGGAATGACAGTGTGATGATGATGATGATGATGATGATGATGATGATGAT 126
Db      566  AAGAAAGGTGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 507
Qy      127  GTCACTGTTGGCGAATGTGTATGATGATGATGATGATGATGATGATGATGATGATGAT 186
Db      506  ATCAGAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 447
Qy      187  CATTTCCCTTACCTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 246
Db      446  CACTTCCCATCTCTTCTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 387
Qy      247  TGTTCCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 306
Db      386  TGTTCCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 327
Qy      307  AAGAGGCGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 366
Db      326  AAGACAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 267
Qy      367  TGTATGACAGCAGCCATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 426
Db      266  TGCATGCTCTTCCGACATTTTCAAGATGATGATGATGATGATGATGATGATGATGATGAT 207
Qy      427  AAAAGCATGACGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 486
Db      206  GACATATGAGAGGTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 147
Qy      487  TTGAGGACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 546
Db      146  CTGCTATCTGTGCAAGGTGCTGGAAGATGATGATGATGATGATGATGATGATGATGATG 87
Qy      547  GAAAAGTGCATTTTCATGTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
Db      86  GAGAAGTGCATTTTCATGTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 33

```

```

RESULT 8
BH714463
LOCUS      BH714463      768 bp      DNA      linear      GSS 20-FEB-2002
DEFINITION      BOMPQ54TF BO_2_3-KB Brassica oleracea genomic clone BOMPQ54, DNA
sequence.
ACCESSION      BH714463
VERSION        BH714463.1 GI:18808105
KEYWORDS      GSS.
SOURCE        Brassica oleracea.
ORGANISM      Brassica oleracea.
REFERENCE      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
AUTHORS      Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
TITLE      Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
JOURNAL      1 (bases 1 to 768)
COMMENT      Town, C.D., Van Aken, S., Uteirback, T. and Fraser, C.M.
              Whole genome shotgun sequencing of Brassica oleracea
              Unpublished (2001)
              Other_GSSs: BOMPQ54TF
              Contact: Chris Town
              TIGR
              9712 Medical Center Drive, Rockville, MD 20850, USA.
              Tel: 301-838-3523
              Fax: 301-838-0208
              Email: cdtown@tigr.org
              DNA is from a doubled haploid provided by Tom Osborn.
              Seq primer: TR

```


Db	490	ACTGCCTTGGCAATCTGTGCGAAGGTGCTGGGAAGAATGTGAGGACAAAGACCTTGTTTA	545
Oy	542	ATTGGGAAGAAGTCATTCATTTGATGGTTCGAGAGGCCATAGTCCTAGGCCACAAGATCTCA	600
Db	550	ATTGGGACAGCTGCATTCATTCATGGTGGAAGAATAGCATTTGTTTGGGTCACAGATATCA	608
RESULT 11			
BH711455/c			
LOCUS			
DEFINITION	BH711455 BOXHT3TfE BO_2_3_KB Brassica oleracea genomic clone BOXHT83, DNA sequence.	854 bp DNA linear GSS 20-FEB-2002	
ACCESSION	BH711455		
VERSION	BH711455.1	GI:18801493	
KEYWORDS	GSS.		
SOURCE	Brassica oleracea		
ORGANISM	Brassica oleracea		
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.		
AUTHORS	1 (bases 1 to 854) Town,C.D., Van Aken,S., Uteirback,T. and Fraser,C.M.		
TITLE	Whole genome shotgun sequencing of Brassica oleracea unpublished (2001)		
JOURNAL	Other_GSS: BOXHT83TR		
COMMENT	Contact: Chris Town		
	TIGR		
	9712 Medical Center Drive, Rockville, MD 20850, USA.		
	Tel: 301-838-3523		
	Fax: 301-838-0208		
	Email: cdtown@tigr.org		
	DNA is from a doubled haploid provided by Tom Osborn.		
	Seq primer: Tff		
	Class: sheared ends.		

Db 188 TTGAGTGCAGGGGAGTACCTACCCATTTCCAGACAGTACTTGGGTGAGCCCGTTTCATG 247
 QY 61 GTTCCCAAGAAAGGTGGAATGACAGTGTACGAGATGAGAGAAATGACTTGATACCA 120
 Db 248 GTTCTTAAGAAAGGTGAGTACATCTGATCACAATAGAGAAAGTGAATCTTCTACC 307
 QY 121 CGAAGTCTACTGTTGGCCAAATGTATGACTATGCGCAAGCTGAATGAAGCCACAGG 180
 Db 308 AGAAGAGTCAAGAGGATAGATGTGATGATGACAGAGAAAGTAACTCAGCCCAAG 367
 QY 181 AAGGACATTTCCCTTACCTTTCATGATGATGATGAGAGAGATGCTGAGAGGAGCA 240
 Db 368 AAGGACCACTTCCCACTTCTTTCATGATGATGATGATGATGATGATGATGATGAT 427
 QY 241 TACTACTGTTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
 Db 428 TACTACTGTTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 487
 QY 301 CAGGAGAAAGCGGCTTACATGATGATGATGATGATGATGATGATGATGATGATGAT 360
 Db 488 CAGGAGAAAGCAATTCACCTGATGATGATGATGATGATGATGATGATGATGATGAT 547
 QY 361 GGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
 Db 548 GGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 607
 QY 421 GTGAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480
 Db 608 ATGAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 667
 QY 481 AGCTGTTGAGAAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 540
 Db 668 GACTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 727
 QY 541 AATTTGAGAAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
 Db 728 AATTTGAGAAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 787

RESULT 14
 BH443545/c
 LOCUS
 DEFINITION
 BH443545 800 bp DNA linear GSS 12-DEC-2001
 BOGG629TF BOGG Brassica oleracea genomic clone BOGG629, DNA
 sequence.
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Brassica oleracea.
 Brassica oleracea
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eustosids II; Brassicales; Brassicaceae; Brassica.
 1 (bases 1 to 800)
 Town, C.D., Van Aken, S., Uterback, T. and Fraser, C.M.
 Whole genome shotgun sequencing of Brassica oleracea
 unpublished (2001)
 COMMENT
 Other_GSSs: BOGG629TF
 Contact: Chris Town
 TIGR
 9712 Medical Center Drive, Rockville, MD 20850, USA.
 Tel: 301-838-3523
 Fax: 301-838-0208
 Email: cdtown@tigr.org
 DNA is from a doubled haploid, provided by Tom Osborn.
 Seq primer: TR
 Class: sheared ends.
 Location/Qualifiers
 1..800
 /organism="Brassica oleracea"
 /strain="T01000DH3"
 /db_xref="taxon:3712"
 /clone="BOGG629"

/clone.lib="BOGG"
 /note="Vector: pPHOS1. site.1: BstXI; 2-3 kb sheared
 genomic DNA inserted into pPHOS1 using BstXI linkers"
 BASE COUNT 238 a 178 c 146 g 238 t
 ORIGIN
 Query Match 50.3% Score 302; DB 17; Length 800;
 Best Local Similarity 70.4% Pred. No. 5.6e-83;
 Matches 404; Conservative 0; Mismatches 170; Indels 0; Gaps 0;
 QY 2 TGGAGGCTGGGCTCATATACCCATCTCTGACAGCGCTTGGGTAAAGCCAGTACAGTGG 61
 Db 575 TAGAGGCTGGGCTCATATATACCCATCTCTGACAGCGCTTGGGTAAAGCCAGTACAGTGG 516
 QY 62 TTCCCAAGAAAGGTGGAATGACAGTGTGATGATGATGATGATGATGATGATGATGAT 121
 Db 515 TTCCCAAGAAAGGTGGAATGACAGTGTGATGATGATGATGATGATGATGATGATGATGAT 456
 QY 122 GAAGTGCATGCTGGTGGGAATGTGTATGCACTATGCAAGCTGAATGAAGCCACAGCA 181
 Db 455 GAAGTGCATGCTGGTGGGAATGTGTATGCACTATGCAAGCTGAATGAAGCCACAGCA 396
 QY 182 AGGACCAATTTCCCTTACCTTTCATGATGATGATGATGATGATGATGATGATGATGAT 241
 Db 395 AGGACCAATTTCCCTTACCTTTCATGATGATGATGATGATGATGATGATGATGATGAT 336
 QY 242 ACTACTGTTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 301
 Db 335 ACTACTGTTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 276
 QY 302 AGGAGAAAGCGGCTTACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 361
 Db 275 AGGAGAAAGCGGCTTACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 216
 QY 362 GGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 421
 Db 215 GGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 156
 QY 422 TGGAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 481
 Db 155 TGGAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 96
 QY 482 GCTGTTGAGAAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 541
 Db 95 TTGTTTAAAAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 36
 QY 542 ATTGGGAAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 575
 Db 35 ACTGGGAAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2

RESULT 15
 BO996483
 LOCUS
 DEFINITION
 BO996483 697 bp mRNA linear EST 22-AUG-2002
 OGG12P10. Yg. ab1 OG_EFGH lettuce serriola Lactuca sativa cDNA clone
 OGG12P10, mRNA sequence.
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Lactuca sativa.
 Lactuca sativa
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Asterales; eusterids II; Asterales; Asteraceae; Lactuceae;
 Lactuca.
 1 (bases 1 to 697)
 Kozik, A., Michelmore, R.W., Knapp, S., Matvienko, M., Riesberg, L.,
 Ito, H., van Damme, M., Lavelle, D., Chevalier, P., Ziegler, J., Ellison,
 P., Kolman, J., Slabaugh, M.S., Livingston, K., Zhou, Y., Lai, Z.,
 Church, S., Jackson, L. and Bradford, K.
 Lettuce and Sunflower ESTs from the Compositae Genome Project
 http://compcomics.ucdavis.edu/
 Unpublished (2002)
 JOURNAL

COMMENT

Contact: Alexander Kozik (R.W.Michelmores)
Department of Vegetable Crops, R.W.Michelmores Lab
University of California at Davis (UCD)
Asmundson Hall, UCD, Davis, CA 95616, USA
Tel: 1-(530)-752-9659
Fax: 1-(530)-752-9659
Email: akozik@ucdavis.org [michelmore@vegmail.ucdavis.edu]
singleton, see http://cgpdb.ucdavis.edu/ for details.
Plate: OG612 row: P column: 10.
Location/Qualifiers

FEATURES

source

1. 697
/organism="Lactuca sativa"
/cultivar="L.seriola"
/db_xref="taxon:4236"
/clone="OG612P10"
/clone_11b="OG-EPGHJ lettuce seriola"
/lab_host="E.coli"
/note="Vector: pBRCDNA5flab. The library was constructed from 10 different sources of RNA from a single genotype. Separate cDNAs were generated using primers that incorporated unique 5' and 3' tags to distinguish each source of RNA. cDNAs were then pooled, size-fractionated, directionally cloned into a custom medium-copy vector and size bias. Details of each source of RNA and library construction can be obtained at http://cgpdb.ucdavis.edu/TAG_LIB-OG-EPGHJ lettuce seriola
TAG_TISSUE=flowers environmental stress
TAG_SEO=CGAATGCGGG"

BASE COUNT 203 a 112 c 162 g 220 t
ORIGIN

Query Match 50.3%; Score 301.6; DB 14; Length 697;
Best Local Similarity 69.6%; Pred. No. 7e-83;

Matches 409; Conservative 0; Mismatches 179; Indels 0; Gaps 0;

```
QY 12 GCTCATATACCCATCTCTACAGACGCTGGGTAAACCCATACAGGTGTTCCAGAA 71
DB 1 GATTATCTACCCGATTCGCGATAGTGTAGTCCAGTGTGTCATGTCCTTAAGAA 60
QY 72 AGGTGAGATGACAGTGTAGAGAGAGATGACTGTATACCAACAGAACTGTAC 131
DB 61 AGGTGAGAGACAGTGTAGAGAGAGATGACTGTATACCAACAGAACTGTAC 120
QY 132 TGGTGGCGATGTATGCACTATGCAAGCTGAATGAACGACGAGAACCATTT 191
DB 121 AGGCTGGAGATTTGTATAGATTACAGAAAGTTAAATACCGCACTAGAGAACCATTT 180
QY 192 CCCCTTACCTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 251
DB 181 CCCATTACCGTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
QY 252 CTGTGATGATATCTCGGATACCAACAGATCGCGGTAGACCCAGAGATCAGAGAAG 311
DB 241 TTGTGATGATATCTGTGATATATATGATTTCTATTCACACAGAAAGATCAACAAAAC 300
QY 312 GGCCTTATACATGCCCTTTGGCGTCTTGTCTTACAGAGAGATTCGGTTATGTA 371
DB 301 TACTTTCACATGTCATTTGGTACCTTTGCTTTAGAGCATGCTTTTGGTTATGTA 360
QY 372 TGCACAGCCACATTTACAGAGTGCATGCTGGCCATTTTTCAGACATGGTGAGAAAAG 431
DB 361 TGCACAGCCACATTTCCAAAGTGATGATGATGATGATGATGATGATGATGATGAT 420
QY 432 CATCAGATATTTTANGAGACATTCGCGTTTGTGACCCCTCATTTGACAGCTGTTGAG 491
DB 421 TGTGAGAGCTTTATGATGATTTTTCAGTCCATGATGATGATGATGATGATGATGAT 480
QY 492 GAACCTAGAGAGCTTCTCAGAGGTGCGAAGAGACTAAGTGTGATGATGATGATGAT 551
DB 481 AAACCTGAGCTGGGATTTGCAAAAGTGTCTCAACATTAATCTGTCTAAATTTGGGAAA 540
QY 552 GTGTCAATTCATGTTGAGAGGGCATAGTCTTAGGCCAACAAGATCTC 599
```

DB 541 ATGCCATTATATGTTAAAGAGGATATTTGTTGGACACAAAGTTTC 588

Search completed: June 20, 2003, 22:39:20
Job time : 2307.77 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2003, 19:31:16 ; Search time 75.6226 Seconds

(without alignments)
6927.165 Million cell updates/sec

Title: US-09-965-553-2

Perfect score: 18

Sequence: 1 tggcgccgtgtgtcgaggga 18

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 segs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:*

1: gb_da:*

2: gb_hcg:*

3: gb_in:*

4: gb_ov:*

5: gb_ov:*

6: gb_ov:*

7: gb_ov:*

8: gb_ov:*

9: gb_ov:*

10: gb_ov:*

11: gb_ov:*

12: gb_ov:*

13: gb_ov:*

14: gb_ov:*

15: gb_ov:*

16: gb_ov:*

17: gb_ov:*

18: gb_ov:*

19: gb_ov:*

20: gb_ov:*

21: gb_ov:*

22: gb_ov:*

23: gb_ov:*

24: gb_ov:*

25: gb_ov:*

26: gb_ov:*

27: gb_ov:*

28: gb_ov:*

29: gb_ov:*

30: gb_ov:*

31: gb_ov:*

32: gb_ov:*

33: gb_ov:*

34: gb_ov:*

35: gb_ov:*

36: gb_ov:*

37: gb_ov:*

38: gb_ov:*

39: gb_ov:*

40: gb_ov:*

41: gb_ov:*

Score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	ID	Description
1	18	100.0	10128	AF186182	Glycine m
2	18	100.0	10510	AF186186	Glycine m
3	18	100.0	83339	AP001311	Arabidops
4	18	100.0	106716	AC006413	Arabidops
5	18	100.0	122116	AP004483	Lotus jap
6	18	100.0	148246	AP003054	Oryza sat
7	18	100.0	189744	AP003577	Oryza sat
8	18	100.0	194640	AP005160	Oryza sat
9	18	100.0	261265	AF427791	Hordeum v
10	17	94.4	67517	AF028277	Hordeum c
11	16.4	91.1	537	AX059280	Sequence
12	16.4	91.1	687	AX059280	Sequence
13	16.4	91.1	1732	BNDNTRNA	B. nigra DNA
14	16.4	91.1	4717	AF439379	Medicago
15	16.4	91.1	6676	SCR1731	X64100 S. cereale D
16	16.4	91.1	8117	TAGLUIAG	X03042 Trilicium ae
17	16.4	91.1	11094	AE005854	AE005854 Caulobact
18	16.4	91.1	12314	PSJ000640	AE007177 Mycobacte
19	16.4	91.1	15085	AE007177	AE006966 Mycobacte
20	16.4	91.1	15417	AE006966	AE006966 Mycobacte
21	16.4	91.1	18789	AF439380	AF439380 Medicago
22	16.4	91.1	33254	AP004537	AD000005 Mycobacte
23	16.4	91.1	36526	MSGY42	284395 Mycobacteri
24	16.4	91.1	36804	MTGCT210	AX059513 Sequence
25	16.4	91.1	48128	AX059513	Sequence
26	16.4	91.1	62916	AB046427	Arabidops
27	16.4	91.1	64165	AC067965	AC067965 Arabidops
28	16.4	91.1	65979	AF254799	Hordeum v
29	16.4	91.1	77132	AP003820	Oryza sat
30	16.4	91.1	77136	AF474982	Hordeum v
31	16.4	91.1	78805	OSJN00123	AL606996 Oryza sat
32	16.4	91.1	80930	AP004894	Lotus jap
33	16.4	91.1	82532	AC006219	Arabidops
34	16.4	91.1	83991	AP004965	Lotus jap
35	16.4	91.1	85748	OSJN00251	AL731602 Oryza sat
36	16.4	91.1	87219	AP002054	Arabidops
37	16.4	91.1	90101	AP003807	Oryza sat
38	16.4	91.1	90627	AP000411	Arabidops
39	16.4	91.1	90840	AB073158	Arabidops
40	16.4	91.1	92281	AP004896	Lotus jap
41	16.4	91.1	94139	AP003805	Oryza sat
42	16.4	91.1	95681	AP002035	Arabidops
43	16.4	91.1	95959	AC006217	Arabidops
44	16.4	91.1	98303	AP004642	Oryza sat
45	16.4	91.1	100595	AP004545	Lotus jap

ALIGNMENTS

RESULT 1

LOCUS AF186182

DEFINITION Glycine max retrovirus-like element Calypso-1, partial sequence.

ACCESSION AF186182

VERSION AF186182.1 GI:6671122

KEYWORDS

SOURCE

ORGANISM Glycine max.

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseolae; Glycine.

REFERENCE 1 (bases 1 to 10128)

AUTHORS Wright,D.A. and Voytas,D.F.

Pred. No. is the number of results predicted by chance to have a

TITLE Calypso: A Heterogeneous Retrovirus-like Element Family from
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 10128)
 AUTHORS Wright, D.A. and Voytas, D.F.
 JOURNAL Direct Submission
 TITLE Submitted (14-SEP-1999) Zoology and Genetics, Iowa State
 UNIVERSITY, 2208 Molecular Biology Bldg., Ames, IA 50011, USA
 FEATURES
 source
 1. 10128
 /organism="Glycine max"
 /strain="L85"
 /db_xref="taxon:3847"
 repeat_region
 /rpt_family="retrovirus-like element Calypso1-1"
 LTR
 primer_bind
 1605..1623
 1759..1717
 /gene="pol"
 /pseudo
 8068..8976
 /gene="envelope-like"
 /pseudo
 BASE COUNT 3031 a 2135 c 2330 g 2632 t
 ORIGIN
 Query Match 100.0%; Score 18; DB 8; Length 10128;
 Best Local Similarity 100.0%; Pred. No. 3.2e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TGGCGCCGTTGTGCGGGA 18
 ||||||||||||||||
 Db 1605 TGGCGCCGTTGTGCGGGA 1622
 RESULT 2
 AF186186 10510 bp DNA linear PLN 10-SEP-2001
 LOCUS Glycine max retrovirus-like element Calypso5-1, partial sequence.
 DEFINITION AF186186
 VERSION AF186186.1 GI:6671126
 KEYWORDS
 SOURCE Glycine max.
 ORGANISM Glycine max.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eustosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
 Glycine.
 1 (bases 1 to 10510)
 Wright, D.A. and Voytas, D.F.
 Calypso: A Heterogeneous Retrovirus-like Element Family from Glycine
 max
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 10510)
 AUTHORS Wright, D.A. and Voytas, D.F.
 JOURNAL Direct Submission
 TITLE Submitted (14-SEP-1999) Zoology and Genetics, Iowa State
 UNIVERSITY, 2208 Molecular Biology Bldg., Ames, IA 50011, USA
 FEATURES
 source
 1. 10510
 /organism="Glycine max"
 /strain="L85"
 /db_xref="taxon:3847"
 31..>10510
 /rpt_family="retrovirus-like element Calypso5-1"
 LTR
 primer_bind
 1478..1495
 1644..8392
 /gene="pol"
 /pseudo
 10144..10510

/gene="envelope-like"
 /pseudo
 BASE COUNT 3186 a 1844 c 2306 g 3174 t
 ORIGIN
 Query Match 100.0%; Score 18; DB 8; Length 10510;
 Best Local Similarity 100.0%; Pred. No. 3.2e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TGGCGCCGTTGTGCGGGA 18
 ||||||||||||||||
 Db 1478 TGGCGCCGTTGTGCGGGA 1495
 RESULT 3
 AP001311/c 83339 bp DNA linear PLN 27-DEC-2000
 LOCUS Arabidopsis thaliana genomic DNA, chromosome 3, P1 clone:MWAll.
 DEFINITION AP001311 BA000014
 ACCESSION AP001311.1 GI:7209747
 VERSION
 KEYWORDS
 SOURCE Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:Mitsui P1
 clone:MWAll.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eustosids II; Brassicales; Brassicaceae; Arabidopsids.
 1 (sites)
 Kaneko, T., Katoh, T., Sato, S., Nakamura, A., Asamizu, E. and Tabata, S.
 Structural analysis of Arabidopsis thaliana chromosome 3. II.
 Sequence features of the 4,251,695 bp regions covered by 90 P1, TAC
 and BAC clones
 DNA Res. 7 (3), 217-221 (2000)
 20363099
 2 (bases 1 to 83339)
 Kaneko, T., Katoh, T., Sato, S., Nakamura, Y., Asamizu, E. and Tabata, S.
 Direct Submission
 Submitted (01-MAR-2000) Yasukazu Nakamura, Kazusa DNA Research
 Institute, Department of Plant Gene Research; 1532-3, Yana,
 Kisarazu, Chiba 292-0812, Japan (E-mail: ynakamu@kazusa.or.jp,
 Tel:81-438-52-3935, Fax:81-438-52-3934)
 Address for correspondence: kaos@kazusa.or.jp
 For the latest information on annotation of this clone, please see
 http://www.kazusa.or.jp/kaos/cgi-bin/agd-graph.cgi?c=MWAll
 Genes with similarity to proteins in the databases are described in
 'product' or 'note' qualifiers. Genes that have no significant
 protein similarity are described as 'unknown protein'.
 The software programs used to predict genes include: Grail
 (Informatics Group, Oak Ridge National Laboratory,
 http://compbio.ornl.gov/Grail-1.3/),
 GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),
 NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of
 Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and
 SplicePredictor (Volker Brendel, Stanford University,
 http://gremmlin.zool.iastate.edu/cgi-bin/sp.cgi).
 Genes encoding tRNAs are predicted by tRNAscan-SE
 (Sean Eddy, Washington University School of Medicine, St. Louis,
 http://genome.wustl.edu/eddy/tRNAscan-SE/).
 This sequence may not be the entire insert of this clone. It may be
 shorter because we remove overlaps between neighboring submissions.
 The 5' clone is T6122 and the 3' clone is MSJ3.
 Location/Qualifiers
 1. 83339
 /organism="Arabidopsis thaliana"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /chromosome="3"
 /clone="MWAll"
 /clone_lib="Mitsui P1"
 complement(join(1709..1889,2017..2172,2316..2444,
 2612..2790))
 /note="gene.id:MWAll.1
 unknown protein"
 CDS

CDS	/codon_start=1 /evidence-not_experimental /protein_id="BAB02224.1" /db_xref="GI:9294327" /translation="MLETKITFEWVEVCOQGFEPIDLPSSQKCRPPDYPLALVKKHPN SGQTVWMSFGKSTLKYSQGMDFVWRIRVPEYPKLILYONIFKLENTQLKEF EYVRLDVQVQSDLSGSFVWLKFRISDILKFWVRIGLFGSGSVQFSGFEMVAV DPLVWLESETMEERLVLCYRDSMMYKERASVKELGVRAEDVYLEMSKNPS" 17169..17450 /note="contains similarity to replication protein A1 gene_id:MVA11.2" /pseudo /codon_start=1 /evidence-not_experimental /join(17769..18220,19125..19210,19322..20811,20886..21233, 21238..21477,21554..21653,21734..22124,22201..22517) /note="contains similarity to unknown protein gb AA06087.1 gene_id:MVA11.3" /pseudo /codon_start=1 /evidence-not_experimental /join(23436..23872,24663..24795,24882..25169,25274..25527, 25580..25913) /note="gene_id:MVA11.4" /pseudo /codon_start=1 /evidence-not_experimental /product="En/Spm-like transposon protein-like" complement(28752..30974) /note="gene_id:MVA11.5" /pseudo /codon_start=1 /evidence-not_experimental /product="retroelement pol polypotein" 38275..38444 /note="gene_id:MVA11.6 unknown protein" /pseudo /codon_start=1 /evidence-not_experimental /join(40447..41304,41458..42397,42583..43109) /note="contains similarity to reverse transcriptase- gene_id:MVA11.7" /pseudo /codon_start=1 /evidence-not_experimental 43611..43908 /note="gene_id:MVA11.8 unknown protein" /pseudo /codon_start=1 /evidence-not_experimental complement(join(44798..44894,44968..45926,46014..46233)) /note="gb AA018631.1 gene_id:MVA11.9 similar to unknown protein" /pseudo /codon_start=1 /evidence-not_experimental complement(join(47335..48223,48359..49241,49413..50566, 50705..51104,51174..51337,51485..52374)) /note="gene_id:MVA11.10" /pseudo /codon_start=1 /evidence-not_experimental /product="gag-protease polypotein-like" complement(join(53267..55857,56051..57537)) /note="contains similarity to pol-like peptide gene_id:MVA11.11" /pseudo /codon_start=1 /evidence-not_experimental	CDS	complement(join(57568..57707,57779..57928,58051..58165, 58276..58305)) /note="contains similarity to disease resistance RPP5 like protein gene_id:MVA11.12" /pseudo /codon_start=1 /evidence-not_experimental join(59311..59421,59522..59671) /note="contains similarity to uroporphyrinogen III methyltransferase gene_id:MVA11.13" /codon_start=1 /evidence-not_experimental /protein_id="BAB02225.1" /db_xref="GI:9294328" /translation="MGFKTLPSPQKIMDHGLFEDPAAVERGTPPLNLYVAELKD FATEIOSAGLVSPPTLIIIGKVELSLPMHCKMSSCLVETQ" complement(63246..66054) /note="gene_id:MVA11.14" /pseudo /codon_start=1 /evidence-not_experimental /product="retroelement pol polypotein" complement(join(70141..70333,70737..70900,71054..71187, 71233..71533,71588..72066,72199..72289,72344..72573)) /note="gene_id:MVA11.15" /pseudo /codon_start=1 /evidence-not_experimental /product="Athlia retroelement protein-like" join(75749..75952,76110..76202,76285..76558,76629..76719, 76801..76930) /note="gb AAD25622.1 gene_id:MVA11.16 strong similarity to unknown protein" /codon_start=1 /evidence-not_experimental /protein_id="BAB02226.1" /db_xref="GI:9294329" /translation="MCIGPFSQKTYEPKPKKTKLQTLDFPNSIDAIGRTKE KKSVAIAPWSPSPFSPEAGPSTVGRILRFEDANIKKQGMGIVLLDEKCS IHAFIPALASHEROVLRIGIIFVNGFVGRCTKXIKTIDHFLPFTATIEVS DVGPIIEREKKFRLRNDNLQALNTIELPDVYGQTFPVGSLNPTSQRLVLRK IDSSVIVISLDDVYATRAHLSGDTTYSVMLITVNRKMG" join(77345..77871,78106..79438) /note="contains similarity to unknown protein gb AAD25621.1 gene_id:MVA11.17" /codon_start=1 /evidence-not_experimental /protein_id="BAB02227.1" /db_xref="GI:9294330" /translation="NEVLDSITKMYTRRRORENIGRIVNLTPTAGDLTYRLILNKVK GATSPDIKTGVGVSHCESKACHAGLIDGKEMDAMDADMAOWMSYLSLSFVLI LITCVSEPKLMSHCHESKADVDLAKQORVNLFPDLKAKLEKTLTLEIFETLQ HEKSLSDVPPEMPQPEKSNKNVPAVSAIAALLLGGRAHWSFKIPINVEDFICD IKIGSLANVLSKVDLIITWDEAPMARHFEAVDRILRLISGDEKALTKLGKIV LLGCDFOILPVIIPORROETVSAAINRSYLSQNMRRQPEBIKFAEWLI QIGDEAPRKTHGIDDOEDNIIIDKNLLPTEMLLEVLCOSVSDPFNFTODELN LKGTAEVAPRNTVEINYLISKVGLAKVEFASIDODEALTEGPEMSVMEVL NSLEPGLPAHRLCLKGVPTMLRLNLRNKEIGCNTRLVTHLGDVKYAKAELISDTT KKRRKVLIPRIILSPDSKHPFTLRROPVPRVCYMTVNRKSGQTLNVALYLPKPV FSHGLYVALSVRTSPKGLTVLDTSKKKGKVTNTVYREVNGLPALITGKKS"	CDS	BASE COUNT 26491 a 15765 c 14391 g 26692 t ORIGIN Query Match 100.0%; Score 18; DB 8; Length 83339; Best Local Similarity 100.0%; Pred. No. 2.4e-02; Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0; OY 1 TGGCGCGTTCGGGGA 18
-----	--	-----	---	-----	---

```

Db          31274   TGGGCGCGTGTGCGGGGA   31257

RESULT 4
AC006413/c  AC006413      106716 bp   DNA      linear   PLN 11-MAR-2002
LOCUS      AC006413
DEFINITION Arabidopsis thaliana chromosome 2 clone F5K7 map ve013, complete
ACCESSION AC006413
VERSION    AC006413
KEYWORDS   GI:20197765
SOURCE     Arabidopsis thaliana
ORGANISM   Arabidopsis thaliana
REFERENCE  1
AUTHORS    Lin,X., Kaul,S., Shea,T.P., Fujii,C.Y., Shen,M., VanKen,S.E.,
            Barnstead,M.E., Mason,T.M., Bowman,C.L., Renning,C.M.,
            Benito,M.-I., Carrera,A.J., Creasy,T.H., Bell,C.R., Town,C.D.,
            Nierman,W.C., Fraser,C.M. and Venter,J.C.
            Unpublished
            2 (bases 1 to 106716)
            Lin,X.
            Direct Submission
            Submitted (09-MAR-2000) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
            3 (bases 1 to 106716)
            Town,C.D. and Kaul,S.
            Direct Submission
            Submitted (27-FEB-2002) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA, cdtonne@igf.org
            On Apr 18, 2002 this sequence version replaced gi:65983560.
            Location/Qualifiers
                1..106716
                /organism="Arabidopsis thaliana"
                /cultivar="Columbia"
                /db_xref="taxon:3702"
                /chromosome="2"
                /map="ve013"
                /clone="F5K7"
                /note="overlap with BAC clone F18P14 (AC006918:1..5067)."
                891..1187
                /gene="At2g06230"
                /note="F5K7.2; predicted by genscan"
                <891..>1187
                /gene="At2g06230"
                891..1187
                /gene="At2g06230"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="AA15238.1"
                /db_xref="GI:20197769"
                /translation="MORIEGVTPMYISLCTRWREMRMEAHPKDMVCGEGRYADIF
                LPLRKVCVEYDSSPIFKPWPRESSDSITGDSGADDFIMRYDDNFKKMPDITLFO"
                complement(1930..4317)
                /gene="At2g06220"
                /note="F5K7.1; predicted by genscan"
                complement(join<1930..2249,2397..2559,2636..2794,
                3087..3184,3509..3698,3901..3995,4056..>4317))
                /gene="At2g06220"
                complement(join(1930..2249,2397..2559,2636..2794,
                3087..3184,3509..3698,3901..3995,4056..4317))
                /gene="At2g06220"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="AA15236.1"
                /db_xref="GI:20197767"
                /translation="MARVRDGTGCHVYESTERPEEVSOTVSEIIEIDDAVAEVP
                TEAPATVSAKPKPKAKTAEAEAEESLDDEASCKTADIEKILFANEGDGELEES
                EBAVNTTKSEVENINCDTCEAEDDDSGPEDEYTCENEDENDEAGNLSLEDEDEDS
                QPLAPETMYSTFEYDNVCKISKIGIVSVLKYERIRSGTDLKLGLVFLVKTLLA

```

[illegible]


```

terminal repeat A (LTR), putative Ty3-type pol
polyprotein-1, putative Ty3-type pol polyprotein-2, long
terminal repeat B (LTR)
complement(17533..18380)
/gene="At2g06190"
/note="F5K7.5"
complement(17533..17932,17938..>18380))
/gene="At2g06190"
complement(join(17533..17932,17938..18380))
/gene="At2g06190"
/codon_start=1
/product="putative Ty3-gypsy-like retroelement pol
polyprotein"
/protein_id="AADI9756.1"
/db_xref="GI:4388718"
/translation="MDFVLGLPTORGVDVFNVDRESKMTHFIACKRTADASNAIK
LEFKEVRLHGVKSTISDRDTFLSFMTSLVMEFGALNRSSPHPIQIDQTEYTN
RLGMMVRSICGDNPKOMDALPOIEEAVSVVVDVLYKPKALGASAEVMAEELV
VKEVYAKLEATGKKKVAADKRRPVEFEGEDVMDVILRGCPKAVGYNKAPKRYG
PKVYRKINDNAVVALPKSMNISNFTFNADIDHEYHADGVLYPEENLRTSSSEVED
VEI"
19174..22307
/gene="At2g06180"
/note="F5K7.6"
join(<19174..19325,19483..19702,19717..20174,20300..20483
20634..20843,20930..21110,21129..21302,21427..21586,
21665..>22307)
/gene="At2g06180"
join(19174..19325,19483..19702,19717..20174,20300..20483,
20634..20843,20930..21110,21129..21302,21427..21586,
21665..22307)
/gene="At2g06180"
/codon_start=1
/product="hypothetical protein"
/protein_id="AADI9757.1"
/db_xref="GI:4388719"
/translation="MTRLSAPHLTAQIRGEEISTISKIYVMAASLNPFHLHSISR
QTLRPSDIDSSONKSNHSTGDDSSSKMSAASROEPPRVTITLGGMPILRPSTRV
DKPKKRAKQKGSLEPTNNDLEVLDSGGDYDPKVAAVYERAGVGGLCRQFLRE
MVAVLPDSGVHPRDALPLRLKLSNVSFRPPRAGVITPGLRGEQIRILRPRAALSG
ODROERPDVLRHLCSSMISDPPRSEGLSGVKLASQILLGGHSHVGDDEPARLTRT
CKNOHNFPRDFTLISVPSKLSNRLATFNPRDVTALPIDPADVITALPAOPDLE
EYPOLYFRSATFOKKNVEYTKSSCSQCSSTASAGDLSAVYAAARTSLASHTOAST
SHIPSPANARPLRRRARLLPVYRLTAELIVIDQEVASKEERKPLAANEAAVSV
DRSSGRSGRGKSGVGRGNRROEQVOTNLGGSGNGRPFMSVYTHDRDRILEDK
AGLAMLRLIRIKGRNCQPGCCCYKQVCSLRATQSGSGFRARRRDPHILCVDERE
VRSAGRVGMLRLKMLSTSKDLGVEYKIKIMQEFGLGDKOMARQRIHLN
EKKDLRSKIVDLTSLAOCYNKAVHAKVFLAASYSKLLSGIKKQVAKKEETVLEIS
AAEVSNNALIDQIVKAAVDLTIEKPRFOAEIDVYEARCTLKEYSDFTLSKIDIEVS
EMLVPRMNVDDQGTPIGLDEFSGSNKDTFPRGLEDPETVATPAGSKKCHG"
complement(23597..26464)
/gene="At2g06170"
/note="F5K7.7"
complement(join(<23597..23821,24081..24417,24700..25436,
25986..26243,26258..>26464))
/gene="At2g06170"
complement(join(23597..23821,24081..24417,24700..25436,
25986..26243,26258..26464))
/gene="At2g06170"

Query Match 100.0%; Score 18; DB 8; Length 106716;
Best Local Similarity 100.0%; Pred. No. 2,3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 TGGCGCGTGTGGGGA 18
Db 49148 TGGCGCGTGTGGGGA 49131

LOCUS AP004483 122116 bp DNA linear PLN 14-DEC-2001
DEFINITION Lotus japonicus genomic DNA, chromosome 1, clone:J173004, TM0016,

```

complete sequence.
AP004483 GI:17736850
AP004483.1 GI:17736850
HNG.
Lotus japonicus DNA, clone_lib=UJT library clone:UJT13004.
Lotus japonicus
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Lotaeae;
Lotus.
1
Sato, S., Kaneko, T., Nakamura, Y., Asamizu, E., Kato, T. and Tabata, S.
Structural Analysis of a Lotus japonicus Genome. I. Sequence
Features and Mapping of Fifty-six TAC Clones which cover the 5.4 Mb
Regions of the Genome
Unpublished
2 (bases 1 to 122116)
Nakamura, Y.
Direct Submission
Submitted (13-DEC-2001) Yasukazu Nakamura, Kazusa DNA Research
Institute, Department of Plant Gene Research; 1532-3, Yana,
Kisarazu, Chiba 292-0812, Japan (E-mail: ynakamu@kazusa.or.jp,
URL: http://www.kazusa.or.jp, Tel: 81-438-52-3935,
Fax: 81-438-52-3934)
Location/Qualifiers
1. 122116
/organism="Lotus japonicus"
/db_xref="taxon:34305"
/chromosome="1"
/clone="UJT13004"
/clone_lib="UJT library"
/note="TAC clone:TM0016"
BASE COUNT 37158 a 23562 c 23033 g 38363 t
ORIGIN
Query Match 100.0%; Score 18; DB 8; Length 122116;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 TGGCGCGGTTGTCGGGGA 18
|||||
Db 52966 TGGCGCGGTTGTCGGGGA 52949
RESULT 6 148246 bp DNA linear PLN 21-MAR-2002
AP003054
LOCUS AP003054
DEFINITION Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1,
PAC clone: P0436D06.
ACCESSION AP003054
VERSION 1
KEYWORDS
SOURCE
ORGANISM
Oryza sativa (japonica cultivar-group) (cultivar:Nipponbare) DNA,
clone:P0436D06.
Oryza sativa (japonica cultivar-group)
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Eumharoidaeae; Oryzeae; Oryza.
1
Sasaki, T., Matsumoto, T. and Yamamoto, K.
Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
clone:P0436D06
Published Only in Database (2000)
2 (bases 1 to 148246)
Sasaki, T., Matsumoto, T. and Yamamoto, K.
Direct Submission
Submitted (20-DEC-2000) Takuji Sasaki, National Institute of
Agrobiological Resources, Rice Genome Research Program; Kannondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail: tsasaki@affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/,
Tel: 81-298-38-7441, Fax: 81-298-38-7469)
On Mar 16, 2001 this sequence version replaced gi:11967926
Genes were predicted from the integrated results of the following:

GENSCAN1.0, BLASTN2.0, BLASTX2.0 as well as SplicePredictor (October 1998 version). The genomic sequence was searched against NCBI Nonredundant Protein database, and the cDNA sequence database at (<http://ncbi.nlm.nih.gov/blast/cdb>) and the NCBI Nonredundant Protein database with BLASTN 2.0. ESTs represent the identified cDNA sequences using BLASTN 2.0 with the corresponding dbi accession no. and RGP clone ID. A gene with identity or significant homology to a protein is classified based on the protein name to indicate the homology level such as same name, 'putative-' and '-like protein'. A gene without significant homology to any protein but with EST homology (covering almost the entire length of partial sequence) is classified as an 'unknown' protein. A gene predicted with a gene prediction program is classified as a 'hypothetical' protein. The orientation of the sequence is from 5' to 3' of the PAC clone. This sequence of P0436D06 clone has an overlap with P0507H06 (DBJ:P003144) clone at 5' end. Detailed information on overlap and assembly quality together with annotation of this entry is available at <http://rgp.dna.affrc.go.jp/GenomeSeq.html>.

FEATURES

source

1..148246 "Oryza sativa (Japonica cultivar-group)"
 /organism="Oryza sativa (Japonica cultivar-group)"
 /cultivar="Nipponbare"
 /db_xref="taxon:39947"
 /chromosome="1"
 /clone="P0436D06"
 join(1618..1860,2195..2404)
 /gene="P0436D06.1"
 join(1618..1860,2195..2404)
 /gene="P0436D06.1"
 /note="hypothetical protein"
 /codon_start=1
 /protein_id="BAB33265.1"
 /db_xref="GI:13359043"
 /translation="MASPCTCRALKMNVSDVSTTLTKRRAPTYCPERSISTIDLPY
 AYHLKQKVITIALAPRPSLRPAVIVPFDNRSAPVNAKQKDWKVAVVQHR
 RHPSTRATRGSDIWRGSOQYRIKNSMAISLDFETIDKSTKSG"
 complement(join(19066..19338,19622..19876,20096..20179,
 21788..22330,25632..25949))
 /gene="P0436D06.2"
 complement(join(19066..19338,19622..19876,20096..20179,
 21788..22330,25632..25949))
 /gene="P0436D06.2"
 complement(join(19066..19338,19622..19876,20096..20179,
 21788..22330,25632..25949))
 /gene="P0436D06.2"
 /note="hypothetical protein
 similar to Arabidopsis thaliana Chromosome 1, F9L1.11"
 /codon_start=1
 /protein_id="BAB33266.1"
 /db_xref="GI:13359044"
 /translation="MGSSDSQAPLPLPRGSHRKEEEEXAAAGKVRGCGGDEGGM
 WREATAEGRSLAPMAIVALLQIMMOLISTVMVGHLEVALAGAAIANSLNVSG
 FSVLMAGLETICGAGAEQYHKLALMYRSIIVLVSVPIAIIWVIFELPL
 IGOPEIASEVGKVALMLPGLEAFVAOCLSKFLOIOLIFPWLSSSTILAFILP
 CFPYKRVGNGAALASVICSICWVEYTVGLIVLSPSCETRAPLTPWEPFSGTST
 RLPAVSAIMLEWMSYELVLISGLIPNPALETSVLISGVSFVSNELGAGNPGA
 RLVAVALSVILCSAVLVSVTLLALRHFIQIAFSNEEVINIVYRMVPLISIVTDS
 LQGVSGIGVSRGCGWQHLAGAYNLGAFIVGVVALFEGFAMHIGGFMGMVAVAG
 ATQVTLISITMTNMRKMGSPFFLHKVLVSQAVSKP"
 join(30567..30579,31283..31494)
 /gene="P0436D06.3"
 join(30567..30579,31283..31494)
 /gene="P0436D06.3"
 /note="hypothetical protein"
 /codon_start=1
 /protein_id="BAB33267.1"
 /db_xref="GI:13359045"
 /translation="MTDCQKPAIPAPPTFTYTMANKTLREPAAPSTDVNAIGPQIMED
 VDFDKSSLITMAQASPFCKPNEDADAHL"
 complement(join(32602..32639,32886..32976))
 /gene="P0436D06.4"
 complement(join(32602..32639,32886..32976))
 /gene="P0436D06.4"
 /note="hypothetical protein"

/codon_start=1
 /protein_id="BAB33268.1"
 /db_xref="GI:13359046"
 /translation="MERGQPCDGLCFRHLGTRGDCPLPAGGVOYGNLGFENVV"
 join(35422..35884,35924..36131,36163..36214)
 /gene="P0436D06.5"
 join(35422..35884,35924..36131,36163..36214)
 /gene="P0436D06.5"
 /note="hypothetical protein"
 /codon_start=1
 /protein_id="BAB33269.1"
 /db_xref="GI:13359047"
 /translation="MHYIMTFNDRKRSGRYSHGNDCCPTREEMYNNAAYRPGQO
 GNNOPPYOAGNNGNFSNQPDLKLVFAQAKTTDALSKRLANDKILDNINIKLG
 FASAPNOISFNKMIETOLALSLVSTNETGTGPGOPSSVEVNAIKATTMGTTG
 KRASDSADKQEPKPTMDEHCDTRLISFPQSRKRPVSDQELARVEVYOKKHINV
 PLLDANTYTRDRSQRRSR"
 complement(36243..37369)
 /note="3' LTR"
 /note="5' LTR"
 /note="5' LTR"
 /note="5' LTR"
 complement(38856..43616)
 /gene="P0436D06.6"
 complement(38856..43616)
 /gene="P0436D06.6"
 /note="probably inactive due to stop codon in CDS
 pseudogene, gag-pol polypeptide"
 /pseudo
 44682..45462
 /note="3' LTR"
 join(46344..48451,48465..49341)
 /gene="P0436D06.7"
 join(46344..48451,48465..48671,48670..49341)
 /gene="P0436D06.7"
 /note="probably inactive due to stop codon in CDS
 pseudogene, retrovirus-related pol polypeptide from
 retrotransposon 17.6"
 /pseudo
 complement(52237..53379)
 /note="5' LTR"
 join(53501..53821,53827..54092,54741..54982,55675..55816,
 55967..56189)
 /gene="P0436D06.8"
 join(53501..53821,53827..54092,54741..54982,55675..55816,
 55967..56189)
 /gene="P0436D06.8"
 /note="hypothetical protein"
 /codon_start=1
 /protein_id="BAB33270.1"
 /db_xref="GI:13359048"
 /translation="MLKDVDPKILFTVLAFTPMRLQADSSVCYTAGIEDVPVKIRD
 FFLVPEVLDMDQTKETPLILRPFSTIGANDVGTGSRFIDGKEVEFEQPT
 EOCSMKRVRESEYSGRSGAODGQPEVNAHFPREGNDAOEPLEDGKTTHTG
 OEARANSEREARLTRKANEVGEKARSTTSRDGRRRLRFKSGSSSHHDES
 STRSSAVSSEDKARRRLNCGDIDIVNREARAYVILSMWEYARREYDELLKTI
 GHILYRELITPLQTEESRAGATRETNMAERKSSSSSTPVOISSVPPSPSE
 WRSSRWELGETTRNMTLDMQGTGLQYMLTEHIAQMQEQATDQAFANINMHQ
 QCGDPL"
 join(57860..57947,57977..58160,58266..58488,58712..58743,
 59094..59199,59234..59612,59620..59747,62210..62476,
 62757..63233,63954..64091)
 /gene="P0436D06.9"
 join(57860..57947,57977..58160,58266..58488,58712..58743,
 59094..59199,59234..59612,59620..59747,62210..62476,
 62757..63233,63954..64091)
 /gene="P0436D06.9"
 /note="hypothetical protein"
 /codon_start=1
 /protein_id="BAB33271.1"
 /db_xref="GI:13359049"
 /translation="MDHCRAEORGCTALOGKTVWSSNNMAGCMGSGMHLVLELA
 IIRARYPKALFSEIGEDRIPSGSKTLPIQTEOSKIGTNOTRPHOIRNGVEGV
 IGRGDIIPARCANDQLSAAGTYNELLPFRPRSPVKGCHCTAEFSTYHAGASANDLWS

IHRHEIAERKAIISGEIFNYPMWLEKEMILTKESSPAPPPLPEPTLADVSVLEFLILY
 TKOCHRCICGAVRGAVLIGFOPSSSSCKRSSLSCAVALCGSGATWATLADY
 VROCTTCACVRCGAPVPAELRLRVLSQCOHARPGCTIDWMLTHSACIVRRRAADAL
 RVRONOPLAMTRORGRHSATCADGSHVGRHREKMRSRKGLTPIEPAPV
 PPPTKTMADTLREFEAPASAEVNAIGPOINMGDVEDLKSLITMAQAPFCGKPED
 ANAHLOFPEICTCTYTKGNFYNGILTPMSRDLDAAGAGAFSKVQAVDLIEKLV
 NMGSEERLOTRORGMHTVEMELFAKLIDLMKCLDHDKROGTVLADSHVCEV
 YGRAPSSPMTHMLVRSTATRVITIGTKARKVRRCICVATTTGTGVHKEVRGTSIAR
 TIKEDILANTLIPMSWYFEKTPQTLLELTTHVHDHTGTLEETPP"
 complement(65302..65898)
 /gene="P0436D06.10"
 complement(65302..65898)
 /gene="P0436D06.10"
 complement(65302..65898)
 /note="hypothetical protein
 similar to Oryza sativa Chromosome 1, P0463F06.26"
 /codon_start=1
 /protein_id="BAB33272.1"
 /db_xref="GI:13359050"
 /translation="MCVDFETDLNKVCPRDHPFLARIDOLVDSMAGCELLSPFDAYS
 HONSMKEDDEKTEITPEGVFTYRMPFGISVGTNOREIOGALGDOIGMNVETI
 DDVYVTKRTGDTLINDLRETPNNLRHKLKINPEKCVSPGSKLLGLVSGGIGIAN
 PEKIKATENKSSTRIKETQRLTGMAHLSFRVARGE"
 complement(join(66849..66972,67147..67435,67625..67669,
 68423..68608,71933..71987,72830..72928))
 /gene="P0436D06.11"
 complement(join(66849..66972,67147..67435,67625..67669,
 68423..68608,71933..71987,72830..72928))

Query Match 100.0%; Score 18; DB 8; Length 148246;
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 TGGGCGCGTGTGCGGGA 18
 Db 37423 TGGGCGCGTGTGCGGGA 37440

RESULT 7 189744 bp DNA linear HTG 21-MAR-2002
 LOCUS AP003577/c
 DEFINITION Oryza sativa (japonica cultivar-group) chromosome 6 clone P0664F03,
 ** SEQUENCING IN PROGRESS ***, in ordered pieces.
 ACCESSION AP003577
 VERSION AP003577.1 GI:13936417
 KEYWORDS HTG; HTGS-PHASE2.
 SOURCE Oryza sativa (japonica cultivar-group) (cultivar: Nipponbare) DNA,
 clone: P0664F03.
 ORGANISM Oryza sativa (japonica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzaceae; Oryza.
 REFERENCE 1
 AUTHORS Sasaki, T., Matsumoto, T. and Yamamoto, K.
 TITLE Oryza sativa nipponbare(GA3) genomic DNA, chromosome 6, PAC
 clone: P0664F03
 JOURNAL Published Only in Database (2001)
 REFERENCE 2 (bases 1 to 189744)
 AUTHORS Sasaki, T., Matsumoto, T. and Yamamoto, K.
 TITLE Direct Submission
 JOURNAL Submitted (02-MAY-2001) Takuji Sasaki, National Institute of
 Agricultural Resources, Rice Genome Research Program, Kannondai
 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
 (E-mail: tsasaki@agr.affrc.go.jp, URL: http://rpg.dna.affrc.go.jp/
 Tel: 81-298-38-7441, Fax: 81-298-38-7468)
 COMMENT NOTE: It currently consists of 1 contigs. Gaps between the contigs
 are represented as runs of N. The order of the pieces is believed
 to be correct as given, however the sizes of the gaps between them
 are based on estimates that have provided by the submitter. This
 sequence will be replaced by the finished sequence as soon as it is
 available and the accession number will be preserved.
 * NOTE: This is a 'working draft' sequence.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and

FEATURES * the accession number will be preserved.
 SOURCE location/Qualifiers
 1..189744
 /organism="Oryza sativa (japonica cultivar-group)"
 /cultivar="Nipponbare"
 /db_xref="taxon:39947"
 /chromosome="6"
 /clone="P0664F03"

BASE COUNT 54659 a 40138 c 40698 g 53849 t 400 others
 ORIGIN

Query Match 100.0%; Score 18; DB 2; Length 189744;
 Best Local Similarity 100.0%; Pred. No. 2.1e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 TGGGCGCGTGTGCGGGA 18
 Db 151774 TGGGCGCGTGTGCGGGA 151757

RESULT 8 194640 bp DNA linear HTG 17-MAY-2002
 LOCUS AP005160/c
 DEFINITION Oryza sativa (japonica cultivar-group) chromosome 7 clone
 OSJNBa0031C24, *** SEQUENCING IN PROGRESS ***, in ordered pieces.
 ACCESSION AP005160
 VERSION AP005160.1 GI:20975286
 KEYWORDS HTG; HTGS-PHASE2.
 SOURCE Oryza sativa (japonica cultivar-group) (cultivar: Nipponbare) DNA,
 clone: OSJNBa0031C24.
 ORGANISM Oryza sativa (japonica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE 1
 AUTHORS Sasaki, T., Matsumoto, T. and Katayose, Y.
 TITLE Oryza sativa nipponbare(GA3) genomic DNA, chromosome 7, BAC
 clone: OSJNBa0031C24
 JOURNAL Published Only in Database (2002)
 REFERENCE 2 (bases 1 to 194640)
 AUTHORS Sasaki, T., Matsumoto, T. and Katayose, Y.
 TITLE Direct Submission
 JOURNAL Submitted (15-MAY-2002) Takuji Sasaki, National Institute of
 Agricultural Sciences, Rice Genome Research Program, Kannondai
 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
 (E-mail: tsasaki@agr.affrc.go.jp, URL: http://rpg.dna.affrc.go.jp/
 Tel: 81-298-38-7441, Fax: 81-298-38-7468)
 COMMENT NOTE: It currently consists of 1 contigs. Gaps between the contigs
 are represented as runs of N. The order of the pieces is believed
 to be correct as given, however the sizes of the gaps between them
 are based on estimates that have provided by the submitter. This
 sequence will be replaced by the finished sequence as soon as it is
 available and the accession number will be preserved.
 * NOTE: This is a 'working draft' sequence.
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.

FEATURES location/Qualifiers
 SOURCE 1..194640
 /organism="Oryza sativa (japonica cultivar-group)"
 /cultivar="Nipponbare"
 /db_xref="taxon:39947"
 /chromosome="7"
 /clone="OSJNBa0031C24"

BASE COUNT 57603 a 39166 c 40961 g 56787 t 103 others
 ORIGIN

Query Match 100.0%; Score 18; DB 2; Length 194640;
 Best Local Similarity 100.0%; Pred. No. 2.1e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 TGGGCGCGTGTGCGGGA 18
 I TGGGCGCGTGTGCGGGA 18


```

gene
  /rpl_type-dispersed
  <29965..>30140
  /gene="ndi"
  /note="coding region not determined; predicted by BLASTn
  and EST search; barley DCINA-induced gene; similar to the
  sequence deposited in GenBank Accession Number AJ250665"
  /evidence-experimental
  31684..32211
  /note="miniature inverted transposable element-like
  sequence; type Hormite3"
  /evidence-not-experimental
  /rpl_family="MITE"
  /rpl_type-dispersed
  32253..32962
  /note="incomplete ORF"
  /transposon="Ac-like transposon"
  34360..34364
  /note="5-bp insertion signature (HORPIA-3)"
  34565..36568
  /note="partial; part of polypeptide and 3' untranslated
  or translated region deleted; the LTRs are intact"
  /evidence-not-experimental
  /rpl_family="HORPIA-3"
  /rpl_type-direct
  /transposon="CopIa/Ty1-like retrotransposon"
  36569..36593
  /note="5-bp insertion signature (HORPIA-3)"
  37785..39765
  /gene="71IN16.9"
  /join(37785..37972,38394..38577,39727..39765)
  /gene="71IN16.9"
  /note="predicted by Genscan; function unknown"
  /codon_start=1
  /product="HY71IN16.9"
  /protein_id="AA22813.1"
  /db_xref="GI:20513852"
  /translation="MPRRPTLSCSVAIRTAAPVPLGRSCMEDGSGAEVWGHDYK
  LVQFPAIDCSVORPFAIVYHLMHGVARVRRRRSRGDIAGALYGVVVVVD
  FRRVGRASIGSPDITDLSGDAAVGCRVMCGT"
  repeat_region
  40068..40310
  /note="Stowaway.3; miniature inverted transposable
  element-like sequence; type Stowaway"
  /rpl_family="MITE"
  /rpl_type-dispersed
  46160..46164
  /note="5-bp insertion signature (BARE-1 SOLOLTR-1.2)"
  46165..47952
  /note="BARE-1 SOLOLTR-1.2"
  /transposon="CopIa/Gypsy-like retrotransposon BARE-1"
  47953..47957
  /note="5-bp insertion signature (BARE-1 SOLOLTR-1.2)"
  49087..49120
  /note="Hormite3.1; miniature inverted transposable
  element-like sequence; type Hormite3"
  /evidence-not-experimental
  /rpl_family="MITE"
  /rpl_type-dispersed
  49126..49163
  /note="Hormite3.2; miniature inverted transposable

Query Match      100.0%; Score 18; DB 8; Length 261265;
Best Local Similarity 100.0%; Pred. NO. 2e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      1 TGGCGCCGTGTCGGGGA 18
        |||||||
Db      256409 TGGCGCCGTGTCGGGGA 256392

RESULT 10
LOCUS      AP002458      67517 bp      DNA      linear      PLN 27-DEC-2000
DEFINITION Arabidopsis thaliana genomic DNA, chromosome 3, BAC clone:122C2.

```

```

ACCESSION      AP002458 BAO00014
KEYWORDS      AP002458.1 GI:8347610
SOURCE      Arabidopsis thaliana (strain:Colombia) DNA, clone_11b:TAMU BAC
              clone:122C2.
ORGANISM      Arabidopsis thaliana
              Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
              Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE      1 (sites)
AUTHORS      Nakamura,Y.
TITLE      Structural Analysis of Arabidopsis thaliana Chromosome 3. III
JOURNAL      Unpublished
REFERENCE      2 (bases 1 to 67517)
AUTHORS      Nakamura,Y.
TITLE      Direct Submission
COMMENT      Submitted (06-JUN-2000) Yasukazu Nakamura, Kazusa DNA Research
              Institute, Department of Plant Gene Research; 1532-3, Yana,
              Kisarazu, Chiba 292-0812, Japan (E-mail:ynakamu@kazusa.or.jp,
              Tel:81-438-52-3935, Fax:81-438-52-3934)
              Address for correspondence: kaos@kazusa.or.jp
              For the latest information on annotation of this clone, please see
              http://www.kazusa.or.jp/kaos/cgi-bin/gdd-graph.cgi?c=122C2
              Genes with similarity to proteins in the databases are described in
              'product' or 'note' qualifiers. Genes that have no significant
              protein similarity are described as 'unknown protein'.
              The software programs used to predict genes include: Grail
              (Informatics Group, Oak Ridge National Laboratory,
              http://compbio.ornl.gov/Grail-1.3/),
              GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),
              NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of
              Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and
              SplicePredictor (Volker Brendel, Stanford University,
              http://gremblin.zool.iastate.edu/cgi-bin/sp.cgi).
              Genes encoding tRNAs are predicted by tRNAscan-SE
              (Sean Eddy, Washington University School of Medicine, St. Louis,
              http://genome.wustl.edu/eddy/tRNAscan-SE/).
              This sequence may not be the entire insert of this clone. It may be
              shorter because we remove overlaps between neighboring submissions.
              The 5' clone is T22B15 and the 3' clone is T1013.
FEATURES
  source
    1..67517
    /organism="Arabidopsis thaliana"
    /strain="Colombia"
    /db_xref="taxon:3702"
    /chromosome="3"
    /clone="122C2"
    /clone_11b="TAMU BAC"
    /complement(32..199)
    /product="(+)delta-cadinene synthase (d-cadinene
    synthase) like"
    /note="CDS is reported in Acc# AP002062
    gene_id:T22B15.28"
    /number=5
    /evidence-not-experimental
    /complement(423..758)
    /product="(+)delta-cadinene synthase (d-cadinene
    synthase) like"
    /note="CDS is reported in Acc# AP002062
    gene_id:T22B15.28"
    /number=4
    /evidence-not-experimental
    /complement(909..1318)
    /product="(+)delta-cadinene synthase (d-cadinene
    synthase) like"
    /note="CDS is reported in Acc# AP002062
    gene_id:T22B15.28"
    /number=3
    /evidence-not-experimental
    /complement(1423..1696)
    /product="(+)delta-cadinene synthase (d-cadinene
    synthase) like"
    /note="CDS is reported in Acc# AP002062
    exon

```

gene_id:"T22B15.28"
/number=2
/evidence-not_experimental
complement(1840..2094)
/product="(+)delta-cadinene synthase (d-cadinene synthase) like"
/note="CDS is reported in Acc# AP002062
gene_id:"T22B15.28"
/number=1
/evidence-not_experimental
complement(join(3258..7313,8240..8346,8681..8834))
/note="gene_id:"T22C2.2"
/pseudo
/codon_start=1
/evidence-not_experimental
/product="non-LTR retroelement reverse transcriptase-like protein"
complement(11186..11911)
/note="gene_id:"T22C2.3"
/codon_start=1
/evidence-not_experimental
/product="Tail non-LTR retroelement protein-like"
/db_xref="GI:877563"
/translation="MADNKRRAVDINLGVDDIPALPEDIVNNAVAENRTLEGRPV
MPRQNLRIIVASMPRIWQSGLVHGRIMEGRQHFETLESELTVDLRGFWAENDW
MILQWMEPIPLPPIFVWQIRGIFPOFNRGVBEHIGALGVLDIDFVAVAR
MGEFARVLHMDITHPLRFRHQRFTAGVNTLRRFRYERLRGCEVCGMLTHDFGACLI
ONGEEOADDDDDDEEHPORTYHN"
complement(join(15703..16226,16332..17328,17506..17657,
17792..17980,18269..18867,18953..19231,20205..20875,
20974..21359))
/note="gene_id:"T22C2.4"
/pseudo
/codon_start=1
/evidence-not_experimental
/product="helicase-like protein"
26176..28509
/note="contains similarity to unknown protein
emb|CAB77997.1
gene_id:"T22C2.5"
/pseudo
/codon_start=1
/evidence-not_experimental
complement(30211..31032)
/note="contains similarity to retroelement pol
polyprotein
gene_id:"T22C2.6"
/pseudo
/codon_start=1
/evidence-not_experimental
33467..33844
/note="emb|CAB81790.1
gene_id:"T22C2.7
similar to unknown protein"
/codon_start=1
/evidence-not_experimental
/protein_id="BA97082.1"
/db_xref="GI:8777564"
/translation="MGADIVADGLTDRPPRLADEPARGLSDRLPEGLSEPPADELVD
KLNLRKPHVDLEDPLFTLMKTRHRLGAAEMTHILFTIPKPREKMLSLGGSLAF
SKTIEYSPDVPCLTLEDLST"
complement(join(34053..35108,35292..35906))
/note="contains similarity to non-LTR retroelement reverse
transcriptase
gene_id:"T22C2.8"
/pseudo
/codon_start=1
/evidence-not_experimental
join(35982..36259,36422..37100)
/note="gene_id:"T22C2.9
pir||T01474
similar to unknown protein"

/pseudo
/codon_start=1
/evidence-not_experimental
complement(join(40707..41698,41792..42666))
/note="gene_id:"T22C2.10"
/pseudo
/codon_start=1
/evidence-not_experimental
/product="retroelement pol
join(44238..44539,44606..45477,45594..45890,45945..46225)
/note="contains similarity to Athalia retroelement ORF1
protein
gene_id:"T22C2.11"
/codon_start=1
/evidence-not_experimental
/protein_id="BA97083.1"
/db_xref="GI:8777565"
/translation="MALDPTASNGEFMRKTEETARKLLENLAGSNHNVYDNRGCG
GSESKQFAELISAKYEQLMRKDQSVNFCEDSSKGVHOFSGDGEDLOAELNFYNGS
TNVENPDQOVYPTQAGSQGQKFPQYQGNKNYQGFQPPAGTGASSFGDEMRLM
QOVLDEQKNNAADINRVDSMYNDLNGKFATLSHVKTLENOVSOIVASMRDHTS
GKYKPKKEQCYAIMIOELREIYVAKOVETNNVYETLVEDKIVEDEPLESPEPPY
VPLKPFGRERIOQROKREYARDEIMQULYRPLQVLHVPSYSLKYLISNRS
IERGVKLISGGEHNAOLVYESQROCKEAOITVNEVLAKEMVNCASIPATIPKKIG
ITNFKPRISILILADRSYQFPMGIAENYHAYGNFYITNPEVLELDEKPHPLNLR
PEVNTVEAIIIDVRSTINLQIGDALIEDLIGTRKNPTIEDCTELSDVEVLDGTDK
AHVILPELOAREDEIEYKVOCKSPKRSKRSITSEMKEGRAVRRVGVKVLRMK
LTMWRPCEGASSRRVH"
join(47271..47464,47537..47661,47751..47818)
/note="gene_id:"T22C2.12
unknown protein"
/codon_start=1
/evidence-not_experimental
/protein_id="BA97084.1"
/db_xref="GI:8777566"
/translation="MSSDNFEFTFASFEDIKSGSLDTNLCVGRYVLYVYRRNO
NDGCPNEMNERFEDYENYEGDKTCLLAKYANDFDNMRHCHDKIILICMRRAKL
KVDGLASQDGAIRYESFTIERSKRS"
join(48668..48804,48893..49148,49230..49299,49414..49567,
49794..49928,50041..50200,50258..50423,50538..50697,
50787..50969)
/note="gb|AA06081.1
gene_id:"T22C2.13
similar to unknown protein"
/pseudo
/codon_start=1
/evidence-not_experimental
join(51608..51745,51834..52098,52211..52313,52391..52622)
/note="gene_id:"T22C2.14"
/codon_start=1
/evidence-not_experimental
/product="replication protein A1-like"

Query Match 94.4% Score 17: DB 8: Length 67517;
Best Local Similarity 100.0% Pred. No. 7.2e+02;
Matches 17: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

QY 1 TGGCGCGGTGTCTCGGG 17
|||||
Db 43810 TGGCGCGGTGTCTCGGG 43826

RESULT 11
AF028277 537 bp DNA linear PLN 13-APR-1999
LOCUS
DEFINITION Hordeum chilense RAPD marker IAS-phc89.
ACCESSION AF028277
VERSION AF028277.1 GI:4581593
KEYWORDS
SOURCE Hordeum chilense.
ORGANISM Hordeum chilense
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

REFERENCE 1 (bases 1 to 537)
 AUTHORS Hernandez, P., Martin, A. and Dorado, G.
 TITLE Direct Submission
 JOURNAL Submitted (06-OCT-1997) Agronomía y Mejora Vegetal, Instituto de Agricultura Sostenible (IAS), Avda. Menéndez Pidal (Alameda del Obispo), Córdoba 14080, Spain
 FEATURES
 SOURCE location/Qualifiers
 1..537
 /organism="Hordeum chilense"
 /strain="H1"
 /db_xref="taxon:15565"
 /clone="IAS-PHC9"

BASE COUNT 150 a 83 c 111 g 193 t

ORIGIN

Query Match 91.1%; Score 16.4; DB 8; Length 537;
 Best Local Similarity 94.4%; Pred. No. 2.7e+03;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGGGCGCGTGTGCGGGA 18
 ||||||||| |||||
 Db 155 TGGGCGCGTGTGCGGGA 172

RESULT 12
 AX059280/c 687 bp DNA linear PAT 17-JAN-2001
 LOCUS AX059280
 DEFINITION Sequence 13 from Patent WO0055325.
 ACCESSION AX059280
 VERSION AX059280.1 GI:12311385
 KEYWORDS
 SOURCE thale cress.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 687)
 Preuss, D., Copenhaver, G. and Keith, K.
 Plant chromosome compositions and methods
 Patent: WO 0055325-A 13 21-SEP-2000;
 The University of Chicago (US)
 FEATURES
 SOURCE 1..687
 /organism="Arabidopsis thaliana"
 /db_xref="taxon:3702"
 BASE COUNT 241 a 135 c 115 g 196 t

ORIGIN

Query Match 91.1%; Score 16.4; DB 6; Length 687;
 Best Local Similarity 94.4%; Pred. No. 2.6e+03;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGGGCGCGTGTGCGGGA 18
 ||||||||| |||||
 Db 22 TGGGCGCGTGTGCGGGA 5

RESULT 13
 BDNATRNA/c 1732 bp DNA linear PLN 06-FEB-1997
 LOCUS BDNATRNA
 DEFINITION B. nigra DNA for tRNA like gene.
 ACCESSION X89901
 VERSION X89901.1 GI:927389
 KEYWORDS Transfer-RNA.
 SOURCE black mustard.
 ORGANISM Brassica nigra
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
 1 (bases 1 to 1732)
 Bhanot, R., Srivastava, P.S., Delseny, M., This, P., Singh Negi, M. and Lakshmi Kumar, M.

TITLE Characterization of species-specific repeats from B. nigra
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 1732)
 AUTHORS Lakshmi Kumar, M.S.
 TITLE Direct Submission
 JOURNAL Submitted (24-JUL-1995) M.S. Lakshmi Kumar, Tata Energy Research Inst., Dabari Seth Block, Habitat Place, Lodhi Road, New Delhi, 110 003, INDIA
 3 (bases 1 to 1732)
 Kapila, R., Negi, M.S., This, P., Delseny, M., Srivastava, P.S. and Lakshmi Kumar, M.
 New family of dispersed repeats from Brassica nigra;
 Characterization and localization
 Theor. Appl. Genet. 930, 1123-1129 (1996)
 FEATURES
 SOURCE location/Qualifiers
 1..1732
 /organism="Brassica nigra"
 /db_xref="taxon:3710"
 /clone.lib="lambda zap II"
 961..1026
 /note="transfer RNA like molecule"
 /note="homology to Bn4 (X67835)"

BASE COUNT 541 a 406 c 339 g 446 t

ORIGIN

Query Match 91.1%; Score 16.4; DB 8; Length 1732;
 Best Local Similarity 94.4%; Pred. No. 2.3e+03;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGGGCGCGTGTGCGGGA 18
 ||||||||| |||||
 Db 1029 TGGGCGCGTGTGCGGGA 1012

RESULT 14
 AF439379 4717 bp mRNA linear PLN 02-MAR-2002
 LOCUS AF439379
 DEFINITION Medicago sativa LTR retroelement MCIRE hypothetical protein mRNA,
 ACCESSION AF439379
 VERSION AF439379
 KEYWORDS complete cds.
 ORIGIN AF439379.1 GI:19071287
 Medicago sativa.
 Medicago sativa
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae; Medicago.
 1 (bases 1 to 4717)
 Ivashuta, S., Naumkina, M., Gau, M., Uchiyama, K., Isobe, S., Mizukami, Y. and Shimanoto, Y.
 Genotype-dependent transcriptional activation of novel repetitive elements during cold acclimation of alfalfa (Medicago sativa L)
 Unpublished
 2 (bases 1 to 4717)
 Ivashuta, S. and Naumkina, M.
 Direct Submission
 Submitted (23-OCT-2001) Graduate School of Agriculture, Hokkaido University, North-9 West-9, Kita-ku, Sapporo, Hokkaido 060-8589, Japan
 FEATURES
 SOURCE location/Qualifiers
 1..4717
 /organism="Medicago sativa"
 /db_xref="taxon:3879"
 /clone="pmc1re7"
 /transposon="LTR retroelement MCIRE"
 /note="putative LTR retroelement"
 1586..2110
 /note="cold-inducible"
 /codon_start=1
 /evidence="not experimental"
 /product="hypothetical protein"

/protein_id="AAL84187.1"
/db_xref="GI:19071288"

/translation="MINMODPTNQRKLNLSLOELIQAQIDNDNANNOSITIGEOENNSN
MFGVGEIVEEVFEQNEVYTLTECCGAPKDELPOERSDSEDEIVDFGEIVKTEKO
ERLLSKKEIFEBQKGRKSKAEIDRVIDEICALFKSLRRTWISHQLYLKFMELPKRR
VSKDVLVSSEFMP"

BASE COUNT 1440 a 754 c 892 g 1631 t
ORIGIN

Query Match 91.1%; Score 16.4; DB 8; Length 4717;
Best Local Similarity 94.4%; Pred. No. 2e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGGCGCGCGTGTGCGGGA 18
|||||
Db 1261 TGGCGCGCGTGTGCGGGA 1278

RESULT 15

LOCUS SCR1731 6676 bp DNA linear PLN 09-SEP-1992
DEFINITION S.cereale DNA for dispersed repeat sequence (R173-1).
ACCESSION X64100
VERSION X64100.1 GI:21196
KEYWORDS dispersed repeat sequence; R173 family; retrotransposon-like.
SOURCE Secale cereale.
ORGANISM Secale cereale.

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Pooidae; Triticeae; Secale.
1 (bases 1 to 6676)
Langridge, P.

REFERENCE
AUTHORS TITLE
JOURNAL Direct Submission
Submitted (09-JAN-1992) P. Langridge, University of Adelaide,
Centre for Cereal Biotechnology, Waite Agricultural Res Inst, Glen
Osmond, South Australia 5064, AUSTRALIA
2 (bases 1 to 6676)
Rogowsky, P.M., Liu, J.Y., Manning, S., Taylor, C. and Langridge, P.
Structural heterogeneity in the R173 family of rye-specific
repetitive DNA sequences
PLANT MOL. BIOL. 20 (1), 95-102 (1992)

JOURNAL MEDLINE 92385773
PUBMED 1325206
COMMENT See also X64100-3.
FEATURES location/Qualifiers
source 1. 6676

/organism="Secale cereale"
/strain="Chinese Spring 1BL/1RS, 1AL/1RS, 1DL/1RS triple
translocation from Rye"
/db_xref="taxon:4550"
/chromosome="1 (short arm)"
/clone_lib="lambda EMBL3"

repeat_unit 225..1021
/note="long terminal repeat 1"

repeat_unit 5325..6121
/note="long terminal repeat 2"

BASE COUNT 1675 a 1374 c 1405 g 2222 t
ORIGIN

Query Match 91.1%; Score 16.4; DB 8; Length 6676;
Best Local Similarity 94.4%; Pred. No. 1.9e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGGCGCGCGTGTGCGGGA 18
|||||
Db 1031 TGGCGCGCGTGTGCGGGA 1048

Search completed: June 20, 2003, 23:23:59
Job time : 78.956 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2003, 18:46:16 ; Search time 10.3019 Seconds
(without alignments)
3934.810 Million cell updates/sec

Title: US-09-965-553-2

Perfect score: 18

Sequence: 1 tggcgccgtgtcgggga 18

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N.GeneSeq.101002.*

- 1: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1980.DAT.*
- 2: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1981.DAT.*
- 3: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1982.DAT.*
- 4: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1983.DAT.*
- 5: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1984.DAT.*
- 6: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1985.DAT.*
- 7: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1986.DAT.*
- 8: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1987.DAT.*
- 9: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1988.DAT.*
- 10: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1989.DAT.*
- 11: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1990.DAT.*
- 12: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1991.DAT.*
- 13: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1992.DAT.*
- 14: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1993.DAT.*
- 15: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1994.DAT.*
- 16: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1995.DAT.*
- 17: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1996.DAT.*
- 18: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1997.DAT.*
- 19: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1998.DAT.*
- 20: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1999.DAT.*
- 21: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA2000.DAT.*
- 22: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA2001A.DAT.*
- 23: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA2001B.DAT.*
- 24: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	18	100.0	18	AAZ35254	Plant retroelement
2	18	100.0	9829	AAZ35271	Soybean retroelement
3	18	100.0	10482	AAZ35275	Soybean retroelement
4	18	100.0	12286	AAZ35261	Plant generic retr
5	16.4	91.1	687	AAZ3107	Arabidopsis thalia
6	16.4	91.1	2364	AAH51978	Mycobacterium tube
7	16.4	91.1	611590	AAZ22303	Arabidopsis thalia
8	16.4	91.1	4403765	AAI99683	Mycobacterium tube
9	16.4	91.1	4403765	AAI99683	Mycobacterium tube

ID	AAZ35254	standard; DNA; 18 BP.	ALIGNMENTS
10	16.4	91.1	4411529 22 AAI99682
11	16.4	91.1	4411529 22 AAI99682
12	15.4	85.6	509 24 ABO35252
13	15.4	85.6	509 24 ABO35253
14	15.4	85.6	778 21 AAF08402
15	15.4	85.6	1649 24 ABK87940
16	15.4	85.6	1759 24 ABN59782
17	15.4	85.6	1770 22 AAH31391
18	15.4	85.6	2181 23 ABL15329
19	15.4	85.6	2657 20 AAK84348
20	15.4	85.6	4165 22 AAH31356
21	15.4	85.6	4358 24 AAI67895
22	15.4	85.6	7984 24 ABK91610
23	15.4	85.6	9280 24 ABK91615
24	15.4	85.6	9285 24 ABK91608
25	15.4	85.6	11765 23 ABL15328
26	15	83.3	3510 24 ABN79856
27	14.8	82.2	27 21 AAZ35424
28	14.8	82.2	128 22 ABA76135
29	14.8	82.2	128 22 ABA40687
30	14.8	82.2	128 22 AAK24801
31	14.8	82.2	128 22 AAK50798
32	14.8	82.2	128 22 AAI27820
33	14.8	82.2	128 22 AAI56792
34	14.8	82.2	128 24 ABS24287
35	14.8	82.2	466 22 AAS29778
36	14.8	82.2	466 22 AAK43860
37	14.8	82.2	510 24 ABN78395
38	14.8	82.2	528 24 ABO29034
39	14.8	82.2	528 24 ABO29035
40	14.8	82.2	561 22 ABA63912
41	14.8	82.2	561 22 ABA31089
42	14.8	82.2	561 22 AAK12418
43	14.8	82.2	561 22 AAK38131
44	14.8	82.2	561 22 AAI18911
45	14.8	82.2	561 22 AAI44032

RESULT 1
AAZ35254

AAZ35254; standard; DNA; 18 BP.

AC AAZ35254;

DT 27-MAR-2000 (first entry)

XX Plant retroelement primer binding site version 2.

DE

XX Retroelement; retrovirus; transgenic plant; gene transfer;

KW primer binding site; soybean; ss.

KW

XX Glycine max.

OS

XX WO9960842-A2.

PN

XX 02-DEC-1999.

PD

XX 28-MAY-1999; 99WO-US11858.

PF

XX 29-MAY-1998; 98US-0087125.

PR

XX 28-MAY-1999; 99US-0087125.

XX

XX (WRIG/) WRIGHT D A.

PA (VOYT/) VOYTAS D F.

XX

XX Wright DA, Voytas DF;

PI

XX WPI; 2000-105586/09.

DR

XX New nucleic acid molecules for imparting agronomically significant

PT characters to plants, especially soybean
 XX
 PS Claim 1(a); Page 72; 118pp; English.
 XX
 CC This oligonucleotide represents a soybean retroelement primer
 CC binding site (version 2). The invention provides molecular tools
 CC in the form of retroelements and retroelement-containing vectors,
 CC cells and plants. Methods are provided for introducing the
 CC retroelements into cells, especially when the retroelement carries
 CC at least 1 agronomically-significant characteristic. In a
 CC preferred method, a helper cell line which expresses gag, pol and
 CC env sequences is used to enable transfer of a secondary construct
 CC which carries an agronomically-significant characteristic and has
 CC retroelement sequences that allow for replication and integration.
 CC Claimed isolated nucleic acid molecules comprise a nucleic acid
 CC sequence selected from a retroelement primer binding site,
 CC envelope, gag, integrase, reverse transcriptase, protease or
 CC RNase-H sequence (see AA235274-61). Also provided are plant
 CC retroviral particles that are used to transfer the nucleic acids
 CC into plant cells.
 XX
 SO Sequence 18 BP; 1 A; 4 C; 9 G; 4 T; 0 other;
 Query Match 100.0%; Score 18; DB 21; Length 18;
 Best Local Similarity 100.0%; Pred. No. 34;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 TGGCGCCGTTGTCGGGA 18
 DB 1 TGGCGCCGTTGTCGGGA 18
 RESULT 2
 AA235271
 ID AA235271 standard; DNA: 9829 BP.
 AC AA235271;
 XX
 DT 27-MAR-2000 (first entry)
 XX
 DE Soybean retroelement Calypso 1-1.
 XX
 KW Retroelement; retrovirus; transgenic plant; gene transfer;
 KM Calypso 1-1; soybean; ss.
 XX
 OS Glycine max.
 XX
 PN WO960842-A2.
 XX
 PD 02-DEC-1999.
 XX
 PF 28-MAY-1999; 99WO-US11858.
 XX
 PR 29-MAY-1998; 98US-0087125.
 PR 28-MAY-1999; 99US-0087125.
 XX
 PA (WRIG/) WRIGHT D A.
 PA (VOYT/) VOYTAS D F.
 PT Wright DA, Voytas DF;
 PI
 DR WPI; 2000-105586/09.
 XX
 PT New nucleic acid molecules for imparting agronomically significant
 PT characters to plants, especially soybean
 XX
 PS Example 3; Page 95-98; 118pp; English.
 XX
 CC This is the nucleotide sequence of the Calypso 1-1 retroelement of
 CC soybean. It was identified by screening of a soybean lambda
 CC library using a reverse transcriptase probe. 2 groups of soybean
 CC retroelements were identified, i.e. Calypso 1-1, 1-2 and 1-3
 CC (see AA235271-73) and Calypso 2-1 and 2-2 (see AA235274-75). The

CC retroelements include gag, pol, env and primer binding site
 CC sequences that can be used in constructs of the invention. The
 CC invention provides molecular tools in the form of retroelements and
 CC retroelement-containing vectors, cells and plants. Methods are
 CC provided for introducing the retroelements into cells, especially
 CC when the retroelement carries at least 1 agronomically-significant
 CC characteristic (ACS). In a preferred method, a helper cell line
 CC which expresses gag, pol and env sequences is used to enable
 CC transfer of a secondary construct which carries an ACS and has
 CC retroelement sequences that allow for replication and integration.
 CC Claimed isolated nucleic acid molecules comprise a nucleic acid
 CC sequence selected from a retroelement primer binding site, envelope,
 CC gag, integrase, reverse transcriptase, protease or RNase-H sequence
 CC (see AA235274-61). Also provided are plant retroviral particles that
 CC particles are used to transfer the nucleic acids into plant cells.
 XX
 SO Sequence 9829 BP; 2948 A; 2067 C; 2260 G; 2554 T; 0 other;
 Query Match 100.0%; Score 18; DB 21; Length 9829;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 TGGCGCCGTTGTCGGGA 18
 DB 1312 TGGCGCCGTTGTCGGGA 1329
 RESULT 3
 AA235275
 ID AA235275 standard; DNA: 10482 BP.
 AC AA235275;
 XX
 DT 27-MAR-2000 (first entry)
 XX
 DE Soybean retroelement Calypso 2-2.
 XX
 KW Retroelement; retrovirus; transgenic plant; gene transfer;
 KM Calypso 2-2; soybean; ss.
 XX
 OS Glycine max.
 XX
 PN WO960842-A2.
 XX
 PD 02-DEC-1999.
 XX
 PF 28-MAY-1999; 99WO-US11858.
 XX
 PR 29-MAY-1998; 98US-0087125.
 PR 28-MAY-1999; 99US-0087125.
 XX
 PA (WRIG/) WRIGHT D A.
 PA (VOYT/) VOYTAS D F.
 PT Wright DA, Voytas DF;
 PI
 DR WPI; 2000-105586/09.
 XX
 PT New nucleic acid molecules for imparting agronomically significant
 PT characters to plants, especially soybean
 XX
 PS Example 3; Page 107-111; 118pp; English.
 XX
 CC This is the nucleotide sequence of the Calypso 2-2 retroelement of
 CC soybean. It was identified by screening of a soybean lambda
 CC library using a reverse transcriptase probe. 2 groups of soybean
 CC retroelements were identified, i.e. Calypso 1-1, 1-2 and 1-3
 CC (see AA235271-73) and Calypso 2-1 and 2-2 (see AA235274-75). The
 CC retroelements include gag, pol, env and primer binding site
 CC sequences that can be used in constructs of the invention. The
 CC invention provides molecular tools in the form of retroelements and
 CC retroelement-containing vectors, cells and plants. Methods are
 CC provided for introducing the retroelements into cells, especially

CC when the retroelement carries at least 1 agronomically-significant
CC characteristic (ACS). In a preferred method, a helper cell line
CC which expresses gag, pol and env sequences is used to enable
CC transfer of a secondary construct which carries an ASC and has
CC retroelement sequences that allow for replication and integration.
CC Claimed isolated nucleic acid molecules comprise a nucleic acid
CC sequence selected from a retroelement primer binding site, envelope,
CC gag, integrase, reverse transcriptase, protease or RNase-H sequence
CC (see AA235254-61). Also provided are plant retroviral particles that
CC particles are used to transfer the nucleic acids into plant cells.
XX
SQ Sequence 10482 BP; 3181 A; 1842 C; 2297 G; 3162 T; 0 other;
Query Match 100.0%; Score 18; DB 21; Length 10482;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TGGCGCCGTTGTCGGGGA 18
|||||
DB 1448 TGGCGCCGTTGTCGGGGA 1465
RESULT 4
AA235261
ID AA235261 standard; DNA; 12286 BP.
XX
AC AA235261;
XX
DT 27-MAR-2000 (first entry)
XX
DE Plant generic retroelement.
XX
KW Retroelement; retrovirus; transgenic plant; gene transfer;
XX soybean; pea; Calypso; athila; cyclops; ss.
XX
OS Glycine max.
OS Arabidopsis thaliana.
OS Pleum sativum.
OS
FH Key Location/Qualifiers
FT CDS 1482..6887
FT /*tag= a
XX
PN WO9960842-A2.
XX
PD 02-DEC-1999.
XX
PF 28-MAY-1999; 99WO-US11858.
XX
PR 29-MAY-1999; 98US-0087125.
PR 28-MAY-1999; 99US-0087125.
XX
PA (WRIG/) WRIGHT D A.
PA (VOYT/) VOYTAS D F.
XX
PI Wright DA, Voytas DF;
XX
DR WPI; 2000-105586/09.
DR P-PSDB; AAY32434.
XX
PT New nucleic acid molecules for imparting agronomically significant
PT characters to plants, especially soybean
XX
PS Claim 1(h); Page 84-88; 118pp; English.
XX
SQ The present sequence comprises a generic plant retroelement
CC obtained from retrovirus-like elements (retroelements) calypso of
CC soybean, cyclops of pea and athila of Arabidopsis thaliana. The
CC invention provides molecular tools in the form of retroelements and
CC retroelement-containing vectors, cells and plants. Methods are
CC provided for introducing the retroelements into cells, especially
CC when the retroelement carries at least 1 agronomically-significant
CC characteristic (ACS). In a preferred method, a helper cell line

CC which expresses gag, pol and env sequences is used to enable
CC transfer of a secondary construct which carries an ASC and has
CC retroelement sequences that allow for replication and integration.
CC Claimed isolated nucleic acid molecules comprise a nucleic acid
CC sequence selected from a retroelement primer binding site,
CC envelope, gag, integrase, reverse transcriptase, protease or
CC RNase-H sequence (see AA235254-61). Also provided are plant
CC retroviral particles that are used to transfer the nucleic acids
CC into plant cells.
XX
SQ Sequence 12286 BP; 3748 A; 2540 C; 2767 G; 3231 T; 0 other;
Query Match 100.0%; Score 18; DB 21; Length 12286;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TGGCGCCGTTGTCGGGGA 18
|||||
DB 1312 TGGCGCCGTTGTCGGGGA 1329
RESULT 5
AAF22107/C
ID AAF22107 standard; DNA; 687 BP.
XX
AC AAF22107;
XX
DT 20-MAR-2001 (first entry)
XX
DE Arabidopsis thaliana centromere conserved sequence 1-2 #13.
XX
XX Centromere; michrosome; vector; ds.
XX
OS Arabidopsis thaliana.
XX
PN WO200055325-A2.
XX
PD 21-SEP-2000.
XX
PF 17-MAR-2000; 2000WO-US07392.
XX
PR 18-MAR-1999; 99US-0125219.
PR 01-APR-1999; 99US-0127409.
PR 18-MAY-1999; 99US-0134770.
PR 13-SEP-1999; 99US-0153584.
PR 17-SEP-1999; 99US-0154603.
XX
PA (UYCH-) UNIV CHICAGO.
XX
PI Preuss D, Copenhaver G, Keith K;
XX
DR WPI; 2000-587529/55.
XX
PT Recombinant DNA construct comprising a plant centromere, useful for
PT producing stably inherited michrosomes which can serve as vectors for
PT the construction of transgenic plant and animal cells
XX
PS Claim 108; Page 274; 1449pp; English.
XX
SQ The present invention relates to a recombinant DNA construct of a plant
CC (Arabidopsis thaliana) centromere. The constructs are useful for
CC producing stably inherited michrosomes which can serve as vectors for
CC the construction of transgenic plant and animal cells expressing
CC selected proteins such as hormones, enzymes, interleukins, clotting
CC factors, cytokines, antibodies, and growth factors.
XX
SQ Sequence 687 BP; 241 A; 135 C; 115 G; 196 T; 0 other;
Query Match 91.1%; Score 16.4; DB 21; Length 687;
Best Local Similarity 94.4%; Pred. No. 1.6e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 TGGCGCCGTTGTCGGGGA 18

```
Db          |||||
            22 TGGCGCGCTGCGGGA 5
RESULT 6
AAH51978/C
ID AAH51978 standard; DNA: 2364 BP.
XX
AC AAH51978;
XX
DE 04-SEP-2001 (first entry)
XX
DE Mycobacterium tuberculosis potential drug target gene SEQ ID 32.
XX
KM Drug target; growth; organism viability; characterisation; ds.
XX
OS Mycobacterium tuberculosis.
XX
PN WO200135317-A1.
XX
PD 17-MAY-2001.
XX
PF 13-NOV-2000; 2000WO-US31152.
XX
PR 12-NOV-1999; 99US-0165086.
PR 12-NOV-1999; 99US-0165124.
PR 01-FEB-2000; 2000US-0179531.
XX
PA (REGC ) UNIV CALIFORNIA.
PI Eisenberg D, Rotstein SH, Marcotte EM;
XX
DR WPI: 2001-329193/34.
DR P-PSDB; AAC81127.
XX
PT Identifying nucleotide or polypeptide sequence for use as drug target,
PT involves providing algorithm that analyzes a functional relationship
PT between nucleotide or polypeptide sequences, and comparing the
PT sequences
XX
PS Disclosure; Page 71-72; 207pp; English.
XX
CC This invention relates to a method for identifying a nucleotide or
CC polypeptide sequence that may be a drug target, or essential for growth
CC or viability of an organism. Polynucleotide sequences AAH51947 - AAH52092
CC represent DNA encoding proteins AAC81096 - AAC81241, Mycobacterium
CC tuberculosis proteins which are potential drug targets. The DNA and
CC protein sequences are used to illustrate the method of the invention. The
CC method involves providing an unknown nucleotide or polypeptide sequences,
CC and comparing it to a number of sequences along with at least one
CC algorithm capable of analyzing a functional relationship between
CC nucleotide and polypeptide sequences. The method is useful for
CC characterizing the function of nucleic acids and polypeptides that may be
CC useful as a target for a drug or essential for the growth or viability of
CC an organism.
XX
SQ Sequence 2364 BP; 467 A; 775 C; 715 G; 407 T; 0 other;
Query Match          91.1%; Score 16.4; DB 22; Length 2364;
Best Local Similarity 94.4%; Pred. No. 1.6e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY      1 TGGCGCGCTGTCGCGGA 18
        |||||
Db      1084 TGGCGCGCTGTCGCGGA 1067
RESULT 7
AAF22303
ID AAF22303 standard; DNA: 611590 BP.
XX
AC AAF22303;
XX
```

```
DT 20-MAR-2001 (first entry)
XX
XX Arabidopsis thaliana chromosome 2 centromere.
XX
XX Centromere; microsome; vector; ds.
XX
XX Arabidopsis thaliana.
XX
OS WO200055325-A2.
XX
PN 21-SEP-2000.
XX
PD 17-MAR-2000; 2000WO-US07392.
XX
PR 18-MAR-1999; 99US-0125219.
PR 01-APR-1999; 99US-0127409.
PR 18-MAY-1999; 99US-0134770.
PR 13-SEP-1999; 99US-0153584.
PR 17-SEP-1999; 99US-0154603.
XX
XX (UYCH-) UNIV CHICAGO.
XX
PI Preuss D, Copenhaver G, Kelch K;
XX
DR WPI: 2000-587529/55.
XX
PT Recombinant DNA construct comprising a plant centromere, useful for
PT producing stably inherited microsome which can serve as vectors for
PT the construction of transgenic plant and animal cells
XX
PS Claim 45; Page 820-959; 1449pp; English.
XX
CC The present invention relates to a recombinant DNA construct of a plant
CC (Arabidopsis thaliana) centromere. The constructs are useful for
CC producing stably inherited microsome which can serve as vectors for
CC the construction of transgenic plant and animal cells expressing
CC selected proteins such as hormones, enzymes, interleukins, clotting
CC factors, cytokines, antibodies, and growth factors.
XX
SQ Sequence 611590 BP; 181893 A; 124460 C; 120254 G; 184983 T; 0 other;
Query Match          91.1%; Score 16.4; DB 21; Length 611590;
Best Local Similarity 94.4%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY      1 TGGCGCGCTGTCGCGGA 18
        |||||
Db      100117 TGGCGACGTGTCGCGGA 100134
RESULT 8
AAI99683
ID AAI99683 standard; DNA: 4403765 BP.
XX
AC AAI99683;
XX
DT 15-JAN-2002 (first entry)
XX
DE Mycobacterium tuberculosis strain H37RV genome SEQ ID NO 2.
XX
KM Mycobacterium tuberculosis; strain H37RV; strain CDC 1551; genome;
KM variation; epidemiology; patient treatment; epidemic monitoring; ds.
XX
OS Mycobacterium tuberculosis.
XX
PN US6294328-B1.
XX
PD 25-SEP-2001.
XX
PF 24-JUN-1998; 98US-0103840.
PR 24-JUN-1998; 98US-0103840.
XX
```

PA (GENO-) INST GENOMIC RES.
XX
PI Fleischmann RD, White OR, Fraser CM, Venter JC;
XX
DR WPI: 2001-647261/74.
XX
PT Evaluating strain variation of Mycobacterium tuberculosis, comprises
PT determining the nucleotide sequence of the strain at positions in the
PT genome corresponding to positions where M. tuberculosis strains CDC
XX 1551 and H37Rv differ -
PS
PS Claim 4; SEQ ID NO 2; 3pp + Sequence Listing; English.
XX
CC The invention relates to evaluating strain variation within and between
CC different populations of the tuberculosis bacterial pathogen,
CC Mycobacterium tuberculosis or related Mycobacterium by determining the
CC nucleotide sequence of the first strain at positions in the complete
CC sequence of the genome that correspond to positions that differ in the
CC nucleotide sequences of M. tuberculosis strains CDC 1551 (AA199683) and
CC H37Rv (AA199682). The method is useful for evaluating strain variation of
CC M. tuberculosis and has valuable application in the fields of
CC tuberculosis genetics, epidemiology, patient treatment and epidemic
CC monitoring.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from USPTO
CC at seqdata.uspto.gov/sequence.html?DocID=6294328B1.
XX
SQ Sequence 4403765 BP; 757105 A; 1447799 C; 1441301 G; 757371 T; 189 other;
XX
Query Match
Best Local Similarity 94.1%; Score 16.4; DB 22; Length 4403765;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 TGGCGCCGTTGTCGGGGA 18
DB 4130872 TGGCGCGCTGTGCGGGA 4130889
II |||||
RESULT 9
AA199683/C
ID AA199683 standard; DNA: 4403765 BP.
XX
XX AA199683;
XX
XX 15-JAN-2002 (first entry)
XX
DE Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 2.
XX
KW Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
XX variation; epidemiology; patient treatment; epidemic monitoring; ds.
XX
OS Mycobacterium tuberculosis.
XX
PN US6294328-B1.
XX
XX 25-SEP-2001.
XX
XX 24-JUN-1998; 98US-0103840.
XX
XX 24-JUN-1998; 98US-0103840.
XX
XX (GENO-) INST GENOMIC RES.
XX
XX Fleischmann RD, White OR, Fraser CM, Venter JC;
XX
XX WPI: 2001-647261/74.
XX
XX Evaluating strain variation of Mycobacterium tuberculosis, comprises
XX determining the nucleotide sequence of the strain at positions in the
XX genome corresponding to positions where M. tuberculosis strains CDC
XX 1551 and H37Rv differ -
XX
XX Claim 4; SEQ ID NO 2; 3pp + Sequence Listing; English.

The invention relates to evaluating strain variation within and between different populations of the tuberculosis bacterial pathogen, Mycobacterium tuberculosis or related Mycobacterium by determining the nucleotide sequence of the first strain at positions in the complete sequence of the genome that correspond to positions that differ in the nucleotide sequences of M. tuberculosis strains CDC 1551 (AA199683) and H37Rv (AA199682). The method is useful for evaluating strain variation of M. tuberculosis and has valuable application in the fields of tuberculosis genetics, epidemiology, patient treatment and epidemic monitoring.
 Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from USPTO at seqlife.uspto.gov/sequence.html?DocID=6294328B1.
 Sequence 4403765 BP; 757105 A; 1447799 C; 1441301 G; 757371 T; 189 other:
 Query Match Best Local Similarity 91.1%; Score 16.4; DB 22; Length 4403765; Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0
 1 TGGCGCCGTTGTGGGGA 18
 |||||
 Db 809593 TGGCGCCGTTGTGGGGA 809576
 |||||
 RESULT 10
 AA199682
 ID AA199682 standard; DNA; 4411529 BP.
 AC AA199682;
 DT 15-JAN-2002 (first entry)
 DE Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 1.
 OS Mycobacterium tuberculosis.
 PN US6294328-B1.
 PD 25-SEP-2001.
 PF 24-JUN-1998; 98US-0103840.
 PR 24-JUN-1998; 98US-0103840.
 PA (GENO-) INST GENOMIC RES.
 PI Fleischmann RD, White OR, Fraser CM, Venter JC;
 DR WPI: 2001-647261/74.
 PT Evaluating strain variation of Mycobacterium tuberculosis, comprises determining the nucleotide sequence of the strain at positions in the genome corresponding to positions where M. tuberculosis strains CDC 1551 and H37Rv differ -
 Claim 3; SEQ ID NO 1; 3pp + Sequence Listing; English.
 The invention relates to evaluating strain variation within and between different populations of the tuberculosis bacterial pathogen, Mycobacterium tuberculosis or related Mycobacterium by determining the nucleotide sequence of the first strain at positions in the complete sequence of the genome that correspond to positions that differ in the nucleotide sequences of M. tuberculosis strains CDC 1551 (AA199683) and H37Rv (AA199682). The method is useful for evaluating strain variation of M. tuberculosis and has valuable application in the fields of tuberculosis genetics, epidemiology, patient treatment and epidemic monitoring.
 Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from USPTO
CC at seqdata.uspto.gov/sequence.html?DocID=6294328B1.
XX
SQ Sequence 4411529 BP; 758565 A; 1449983 C; 1444602 G; 758379 T; 0 other;
Query Match 91.1%; Score 16.4; DB 22; Length 4411529;
Best Local Similarity 94.4%; Pred. No. 1.1e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 TGGCGCCGTTGCGGGA 18
DB 4138678 TGGCGCCGTTGCGGGA 4138695
RESULT 11
ID AA199682 standard; DNA; 4411529 BP.
XX
AC AA199682;
XX
DT 15-JAN-2002 (first entry)
DE Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 1.
XX
KW Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
KM variation; epidemiology; patient treatment; epidemic monitoring; ds.
XX
OS Mycobacterium tuberculosis.
XX
PN US6294328-B1.
XX
PD 25-SEP-2001.
XX
PE 24-JUN-1998; 98US-0103840.
XX
PR 24-JUN-1998; 98US-0103840.
XX
PA (GENO-) INST GENOMIC RES.
XX
PI Fleischmann RD, White OR, Fraser CM, Venter JC;
XX
DR WPI: 2001-647261/74.
XX
PT Evaluating strain variation of Mycobacterium tuberculosis, comprises
PT determining the nucleotide sequence of the strain at positions in the
PT genome corresponding to positions where M. tuberculosis strains CDC
PT 1551 and H37Rv differ -
XX
PS Claim 3; SEQ ID NO 1; 3pp + Sequence Listing; English.
XX
CC The invention relates to evaluating strain variation within and between
CC different populations of the tuberculosis bacterial pathogen,
CC Mycobacterium tuberculosis or related Mycobacterium by determining the
CC nucleotide sequence of the first strain at positions in the complete
CC sequence of the genome that correspond to positions that differ in the
CC nucleotide sequences of M. tuberculosis strains CDC 1551 (AA199683) and
CC H37Rv (AA199682). The method is useful for evaluating strain variation of
CC M. tuberculosis and has valuable application in the fields of
CC tuberculosis genetics, epidemiology, patient treatment and epidemic
CC monitoring.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from USPTO
CC at seqdata.uspto.gov/sequence.html?DocID=6294328B1.
XX
SQ Sequence 4411529 BP; 758565 A; 1449983 C; 1444602 G; 758379 T; 0 other;
Query Match 91.1%; Score 16.4; DB 22; Length 4411529;
Best Local Similarity 94.4%; Pred. No. 1.1e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 TGGCGCCGTTGCGGGA 18
DB 807416 TGGCGCCGTTGCGGGA 807399

RESULT 12
ID ABQ35252 standard; DNA; 509 BP.
XX
AC ABQ35252;
XX
DT 12-JUL-2002 (first entry)
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 21843.
XX
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
OS Homo sapiens.
XX
PN WO200218632-A2.
XX
PD 07-MAR-2002.
XX
PF 01-SEP-2001; 2001WO-EP10074.
XX
PR 01-SEP-2000; 2000DE-1043826.
XX
PR 05-SEP-2000; 2000DE-1044543.
XX
PA (EPiG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K, Guelig D;
XX
DR WPI: 2002-371829/40.
XX
PT Determining the degree of cytosine methylation in genomic DNA, useful
PT for diagnosis and prognosis, comprises selective hybridization of
PT amplicons from chemically treated DNA -
XX
PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one
CC member of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridisation to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridised to the two
CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (i) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP's) and (ii) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC ABQ3410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.
XX
SQ Sequence 509 BP; 49 A; 63 C; 174 G; 223 T; 0 other;
Query Match 85.6%; Score 15.4; DB 24; Length 509;
Best Local Similarity 94.1%; Pred. No. 4.8e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 2 GCGCGCGTTGCGGGA 18
DB 207 GCGCGCGTTGCGGGA 223
RESULT 13

ABQ35253/c
 ID ABQ35253 standard; DNA; 509 BP.
 AC ABQ35253;
 XX
 DT 12-JUL-2002 (first entry)
 XX
 DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 21844.
 XX
 KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
 KW drug; side effect; cancer; central nervous system; cardiovascular;
 KW gastrointestinal; respiratory system; single nucleotide polymorphism;
 KW SNP; cell differentiation; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200218662-A2.
 XX
 PD 07-MAR-2002.
 XX
 PF 01-SEP-2001; 2001WO-EPI0074.
 XX
 PR 01-SEP-2000; 2000DE-1043826.
 PR 05-SEP-2000; 2000DE-1044543.
 XX
 PA (EPIC-) EPICENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K, Guetig D;
 DR WPI: 2002-371829/40.
 XX
 PT Determining the degree of cytosine methylation in genomic DNA, useful
 PT for diagnosis and prognosis, comprises selective hybridization of
 PT amplicons from chemically treated DNA
 XX
 PS Claim 12; 56bp + Sequence Listing; 56bp; German.
 XX
 CC This invention describes a novel method for determining the degree of
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
 CC genomic sample of DNA. The sample is treated chemically to convert
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic
 CC DNA that contains the target C is amplified to form a labeled amplicon.
 CC The amplicon is hybridised to two classes, each with at least one
 CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
 CC and the degree of hybridisation to both classes is determined from the
 CC label on the amplicon. From the ratio of labels hybridised to the two
 CC classes of oligomers, the degree of methylation is calculated. The method
 CC is used: (i) for diagnosis and/or prognosis of side effects of
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory
 CC systems etc., particularly by detecting mutations or single nucleotide
 CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
 CC types and for investigating cell differentiation. The method allows the
 CC methylation status of many C residues to be determined simultaneously.
 CC ABQ3410-ABQ4121 represent genomic DNA sequences used to illustrate the
 CC method for determining the degree of cytosine methylation described in
 CC the disclosure of the invention.
 XX
 SO Sequence 509 BP; 223 A; 174 C; 63 G; 49 T; 0 other;

Query Match 85.6%; Score 15.4; DB 24; Length 509;
 Best Local Similarity 94.1%; Pred. No. 4.8e+02;
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 GCGCCCGTTGTCGGGGA 18
 ||||| ||||| ||||| |||||
 DB 303 GCGCCGTTGTCGGGGA 287

RESULT 14
 AAF08402/c
 ID AAF08402 standard; CDNA; 778 BP.
 XX

AC AAF08402;
 XX
 DT 13-MAR-2001 (first entry)
 XX
 DE Fusarium venenatum EST SEQ ID NO:925.
 XX
 KW Multiple gene expression; filamentous fungal cell; EST;
 KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
 KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
 KW culture condition; environmental stress; spore morphogenesis;
 KW metabolic pathway engineering; catabolic pathway engineering; ss.
 XX
 OS Fusarium venenatum.
 XX
 PN WO200056762-A2.
 XX
 PD 28-SEP-2000.
 XX
 PF 22-MAR-2000; 2000WO-US07781.
 XX
 PR 22-MAR-1999; 99US-0273623.
 XX
 PA (NOVO) NOVO NORDISK BIOTECH INC.
 PA (NOVO) NOVO NORDISK AS.
 XX
 PI Berka RM, Rey MM, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
 DR WPI: 2000-594572/56.
 XX
 PT Monitoring differential expression of genes in filamentous fungal cells
 PT uses fluorescence-labeled nucleic acids isolated from the cells and a
 PT substrate of expressed sequence tags -
 XX
 PS Claim 86; Page 737; 3161bp; English.
 XX
 CC The present invention describes a method for monitoring differential
 CC expression of genes in a first filamentous fungal (FF) cell relative to
 CC expression of the same genes in one or more second filamentous fungal
 CC cells. The method uses fluorescence-labeled nucleic acids isolated from
 CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
 CC are used in the methods for monitoring differential expression of genes
 CC in a first filamentous fungal (FF) cell relative to expression of the
 CC same genes in one or more second filamentous fungal cells. Monitoring
 CC the global expression of genes from FF cells allows the production
 CC potential of the microorganisms to be improved. New genes may be
 CC discovered, possible functions of unknown open reading frames can be
 CC identified and gene copy number variation and stability can be
 CC monitored. The expression of genes can be used to study how FF cells
 CC adapt to changes in culture conditions, environmental stress, spore
 CC morphogenesis, recombination, metabolic or catabolic pathway
 CC engineering. Using ESTs provides several advantages over genomic or
 CC random cDNA clones including elimination of redundancy as one spot on an
 CC array equals one gene or open reading frame, and organisation of the
 CC microarrays based on function of the gene products to facilitate
 CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
 CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
 CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
 CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
 CC all specifically claimed in the present invention.
 XX
 SO Sequence 778 BP; 184 A; 232 C; 184 G; 176 T; 2 other;

Query Match 85.6%; Score 15.4; DB 21; Length 778;
 Best Local Similarity 94.1%; Pred. No. 4.7e+02;
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TGGCGCCGTTGTCGGGG 17
 ||||| ||||| ||||| |||||
 DB 523 TGGCGCCGTTGTCGGGG 507

RESULT 15
 AAF087940
 ABK87940

ID ABR87940 standard; cDNA; 1649 BP.
 AC ABR87940;
 DT 07-OCT-2002 (first entry)
 DE Human short-chain dehydrogenase/reductase, 25206, cDNA.
 KM Human; gene: ss; cytosolic; noctropic; neuroprotective; antitumour;
 KM 25206; short-chain dehydrogenase; short-chain reductase; SDR; retinol;
 KM retinal; retinoic acid; embryonic development; spermatogenesis;
 KM epithelial differentiation; metabolism; development; growth;
 KM proliferative disorder; haematopoietic disorder; cancer; tumour;
 KM differentiative disorder; carcinoma; sarcoma; Hodgkin's disease;
 KM neurodegenerative disorder; Alzheimer's disease; Parkinsonism;
 KM progressive supranuclear palsy; reproductive disorder; ovarian tumours;
 KM polycystic ovarian disease.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT CDS 213..1073
 FT /*tag= a
 FT /product= "25206"
 FT /note= "This CDS is specifically claimed in claim 1"
 PN MO200244356-A2.
 PD 06-JUN-2002.
 PE 29-NOV-2001; 2001MO-US45040.
 PF 30-NOV-2000; 2000US-250186P.
 PR (MILL-) MILLENIUM PHARM INC.
 PA Meyers RE, Macbeth KJ;
 PI WPI: 2002-547697/58.
 DR P-PSDB; AAU99344.
 DX Novel isolated 25206 polypeptide, a human short-chain
 PT dehydrogenase/reductase family member, useful as diagnostic targets and
 PT therapeutic agents for treating cancer, Alzheimer's disease,
 PT Parkinsonism, ovarian tumors
 XX
 PS Claim 1: Page 101-102; 117pp; English.
 XX
 CC The invention discloses an isolated 25206 polypeptide, a human
 CC short-chain dehydrogenase/reductase (SDR) family member. Members of this
 CC family catalyse the reversible, rate limiting conversion of retinol to
 CC retinal. Retinal is then converted to retinoic acid which plays a key
 CC role in the regulation of embryonic development, spermatogenesis and
 CC epithelial differentiation. Short chain dehydrogenases are important in
 CC the metabolism of small molecules, production/removal of biologically
 CC important molecules that modulate development and growth, elimination of
 CC toxins and associated physiological processes and pathological
 CC conditions. The polynucleotide and polypeptide are useful for identifying
 CC compounds which binds to them, for identifying compounds which modulate
 CC the activity of the polypeptide, raising antibodies, for detecting the
 CC presence of the polypeptide in a sample, for modulating aberrant activity
 CC of 25206-expressing cells (e.g. a cancerous, pre-cancerous or neural
 CC cell) and for treating or preventing disorders characterised by the
 CC aberrant activity. The 25206 molecules can act as diagnostic targets and
 CC therapeutic agents for controlling one or more cellular proliferative,
 CC including haematopoietic disorders, or differentiative disorders (e.g.
 CC cancer, tumours, carcinoma, sarcoma and Hodgkin's disease), neural (e.g.
 CC neurodegenerative disorders including Alzheimer's disease, Parkinsonism
 CC and progressive supranuclear palsy) and reproductive disorders (e.g.
 CC polycystic ovarian disease and ovarian tumours). The sequence presented
 CC is the human short-chain dehydrogenase/reductase, 25206, cDNA.
 XX
 SQ Sequence 1649 BP; 275 A; 543 C; 532 G; 299 T; 0 other;

Query Match: 85.6%; Score 15.4; DB 24; Length 1649;
 Best Local Similarity 94.1%; Pred. No. 4.6e+02;
 Matches 16; Conservative 1; Indels 0; Gaps 0;
 QY 2 GCGCCGTTGTGCGGGA 18
 ||||| |||||
 Db 1274 GCGCCCTTGTGCGGGA 1290

Search completed: June 20, 2003, 21:58:31
 Job time : 22.3019.secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2003, 21:44:01 ; Search time 2.17925 Seconds

(without alignments)
2533.070 Million cell updates/sec

Title: US-09-965-553-2
Perfect score: 18
Sequence: 1 tggcgcctgttcgggga 18

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 segs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

Issued Patents_NA: *
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq: *
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq: *
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq: *
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq: *
5: /cgn2_6/ptodata/1/ina/PCROS.COMB.seq: *
6: /cgn2_6/ptodata/1/ina/Backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Query	Match	Length	ID	Description
	1	18	100.0	18	4	US-09-322-478-2	Sequence 2, Appl1
	2	18	100.0	27	4	US-09-322-478-39	Sequence 39, Appl1
	3	18	100.0	9829	4	US-09-322-478-19	Sequence 19, Appl1
	4	18	100.0	10482	4	US-09-322-478-23	Sequence 23, Appl1
	5	18	100.0	12286	4	US-09-322-478-17	Sequence 17, Appl1
	6	16.4	91.1	4403765	4	US-09-103-840A-2	Sequence 2, Appl1
	7	16.4	91.1	4403765	4	US-09-103-840A-2	Sequence 2, Appl1
	8	16.4	91.1	4411529	4	US-09-103-840A-1	Sequence 1, Appl1
	9	16.4	91.1	4411529	4	US-09-103-840A-1	Sequence 1, Appl1
	10	14.8	82.2	1018	4	US-09-452-239-35	Sequence 35, Appl1
	11	14.8	82.2	1078	4	US-09-452-239-41	Sequence 41, Appl1
	12	14.8	82.2	2409	4	US-09-230-225B-3	Sequence 3, Appl1
	13	14.8	82.2	3348	4	US-09-302-620B-94	Sequence 94, Appl1
	14	14.8	82.2	6909	4	US-09-199-637A-111	Sequence 111, Appl1
	15	14.4	80.0	231	4	US-08-611-757-14	Sequence 11, Appl1
	16	14.4	80.0	231	5	PCR-US95-05980-14	Sequence 14, Appl1
	17	14.4	80.0	1851	4	US-08-867-611-29	Sequence 29, Appl1
	18	14.4	80.0	1851	4	PCR-US92-06965A-34	Sequence 34, Appl1
	19	13.8	76.7	41	1	US-08-530-492-143	Sequence 143, Appl1
	20	13.8	76.7	41	1	US-08-906-517-143	Sequence 143, Appl1
	21	13.8	76.7	273	4	US-09-068-101-3	Sequence 3, Appl1
	22	13.8	76.7	462	4	US-09-199-637A-212	Sequence 212, Appl1
	23	13.8	76.7	462	4	US-09-199-637A-210	Sequence 210, Appl1
	24	13.8	76.7	624	4	US-09-199-637A-210	Sequence 210, Appl1
	25	13.8	76.7	750	3	US-09-167-717-3	Sequence 3, Appl1
	26	13.8	76.7	1271	4	US-09-231-227-1	Sequence 1, Appl1
	27	13.8	76.7	1386	1	US-08-672-571A-4	Sequence 4, Appl1

C	28	13.8	76.7	1413	4	US-08-984-709A-52	Sequence 52, Appl1
C	29	13.8	76.7	1425	2	US-08-356-060A-6	Sequence 6, Appl1
C	30	13.8	76.7	1425	4	US-08-460-900C-6	Sequence 6, Appl1
C	31	13.8	76.7	1425	4	US-08-674-509B-6	Sequence 6, Appl1
C	32	13.8	76.7	1425	4	US-08-954-698-6	Sequence 6, Appl1
C	33	13.8	76.7	1425	4	US-08-957-874-6	Sequence 6, Appl1
C	34	13.8	76.7	1425	4	US-09-325-256-10	Sequence 10, Appl1
C	35	13.8	76.7	1473	1	US-08-672-571A-2	Sequence 2, Appl1
C	36	13.8	76.7	1576	1	US-08-748-591-5	Sequence 5, Appl1
C	37	13.8	76.7	1576	1	US-08-748-591-10	Sequence 10, Appl1
C	38	13.8	76.7	1669	4	US-08-984-709A-51	Sequence 51, Appl1
C	39	13.8	76.7	1931	1	US-08-530-492-2	Sequence 2, Appl1
C	40	13.8	76.7	1931	4	US-08-906-517-2	Sequence 2, Appl1
C	41	13.8	76.7	2081	4	US-09-149-476-123	Sequence 123, Appl1
C	42	13.8	76.7	2414	6	5248599-1	Patent No. 5248599
C	43	13.8	76.7	2436	4	US-09-199-637A-92	Sequence 92, Appl1
C	44	13.8	76.7	2572	4	US-09-221-017B-225	Sequence 225, Appl1
C	45	13.8	76.7	2804	1	US-08-446-794A-3	Sequence 3, Appl1

ALIGNMENTS

```
RESULT 1
US-09-322-478-2
; Sequence 2, Application US/09322478
; Patent No. 6331662
; GENERAL INFORMATION:
; APPLICANT: Wright, David A.
; TITLE OF INVENTION: Plant Retroelements and Methods Related Thereto
; FILE REFERENCE: P-1065 ISURF Plant Retroelement
; CURRENT FILING DATE: 1999-05-28
; EARLIER FILING DATE: 1998-05-29
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 18
; TYPE: DNA
; ORGANISM: Glycine max
US-09-322-478-2

Query Match          100.0%; Score 18; DB 4; Length 18;
Best Local Similarity 100.0%; Pred.No. 5.8;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TGGCCCGTGTTCGGGGA 18
        |||||
DB       1 TGGCCCGTGTTCGGGGA 18

RESULT 2
US-09-322-478-39
; Sequence 39, Application US/09322478
; Patent No. 6331662
; GENERAL INFORMATION:
; APPLICANT: Wright, David A.
; TITLE OF INVENTION: Plant Retroelements and Methods Related Thereto
; FILE REFERENCE: P-1065 ISURF Plant Retroelement
; CURRENT FILING DATE: 1999-05-28
; EARLIER FILING DATE: 1998-05-29
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 39
; LENGTH: 27
; TYPE: DNA
; ORGANISM: Glycine max
US-09-322-478-39
```

Query Match 100.0%; Score 18; DB 4; Length 27;
Best Local Similarity 100.0%; Pred. No. 5.8;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGCGCCGTTGTGCGGGA 18
|||||
Db 5 TGGCGCCGTTGTGCGGGA 22

RESULT 3
US-09-322-478-19
; Sequence 19, Application US/09322478
; Patent No. 6331662
; GENERAL INFORMATION:

; APPLICANT: Wright, David A.
; TITLE OF INVENTION: Plant Retroelements and Methods Related Thereto
; FILE REFERENCE: P-1065 ISURF Plant Retroelement
; CURRENT APPLICATION NUMBER: US/09/322,478
; EARLIER FILING DATE: 1999-05-28
; EARLIER APPLICATION NUMBER: 60/087125
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 19
; LENGTH: 9829
; TYPE: DNA
; ORGANISM: Glycine max

US-09-322-478-19

Query Match 100.0%; Score 18; DB 4; Length 9829;
Best Local Similarity 100.0%; Pred. No. 4.8;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGCGCCGTTGTGCGGGA 18
|||||
Db 1312 TGGCGCCGTTGTGCGGGA 1329

RESULT 4
US-09-322-478-23

; Sequence 23, Application US/09322478
; Patent No. 6331662
; GENERAL INFORMATION:
; APPLICANT: Wright, David A.
; TITLE OF INVENTION: Plant Retroelements and Methods Related Thereto
; FILE REFERENCE: P-1065 ISURF Plant Retroelement
; CURRENT APPLICATION NUMBER: US/09/322,478
; EARLIER FILING DATE: 1999-05-28
; EARLIER APPLICATION NUMBER: 60/087125
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 23
; LENGTH: 10482
; TYPE: DNA
; ORGANISM: Glycine max
US-09-322-478-23

Query Match 100.0%; Score 18; DB 4; Length 10482;
Best Local Similarity 100.0%; Pred. No. 4.8;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGCGCCGTTGTGCGGGA 18
|||||
Db 1448 TGGCGCCGTTGTGCGGGA 1465

RESULT 5
US-09-322-478-17
; Sequence 17, Application US/09322478

; Patent No. 6331662
; GENERAL INFORMATION:
; APPLICANT: Wright, David A.
; APPLICANT: Voytas, Daniel F.
; TITLE OF INVENTION: Plant Retroelements and Methods Related Thereto
; FILE REFERENCE: P-1065 ISURF Plant Retroelement
; CURRENT APPLICATION NUMBER: US/09/322,478
; EARLIER FILING DATE: 1999-05-28
; EARLIER APPLICATION NUMBER: 60/087125
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 17
; LENGTH: 12286
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: plant
US-09-322-478-17

Query Match 100.0%; Score 18; DB 4; Length 12286;
Best Local Similarity 100.0%; Pred. No. 4.8;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGCGCCGTTGTGCGGGA 18
|||||
Db 1312 TGGCGCCGTTGTGCGGGA 1329

RESULT 6

US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLETSCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 91.1%; Score 16.4; DB 4; Length 4403765;
Best Local Similarity 94.4%; Pred. No. 18;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGGCGCCGTTGTGCGGGA 18
|||||
Db 4130872 TGGCGCCGTTGTGCGGGA 4130889

RESULT 7

US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLETSCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.

```

: APPLICANT VENTER, John C.
: TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
: TITLE OF INVENTION: TUBERCULOSIS
: TITLE OF INVENTION: TUBERCULOSIS
: FILE REFERENCE: 24366-20007.00
: CURRENT APPLICATION NUMBER: US/09/103,840A
: CURRENT FILING DATE: 1998-06-24
: NUMBER OF SEQ ID NOS: 2
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 2
: LENGTH: 4403765
: TYPE: DNA
: ORGANISM: Mycobacterium tuberculosis
: FEATURE:
: OTHER INFORMATION: CDC 1551
: OTHER INFORMATION: "n" bases at various positions throughout the sequence
: OTHER INFORMATION: represent a, t, c or g
: US-09-103-840A-2

```

Query Match	91.1%	Score 16.4	DB 4	Length 4403765
Best Local Similarity	94.4%	Pred. No. 18		
Matches 17; Conservative	0	Mismatches 1	Indels 0	Gaps 0
QY	1	TGGCGCCGTTCTCGGGCA	18	
Db	809593	TGGCGCCGTTCTCGGACA	809576	

```

US-09-103-840A-1
Sequence 1, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE REFERENCE: 24366-20007_00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 441529
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
OTHER INFORMATION: H37Rv
US-09-103-840A-1

```

	Query Match	91.1%	Score 16.4;	DB 4,	Length 4411529;
	Best Local Similarity	94.4%;	Pred. NO. 18;		
	Matches	17;	Conservative	0;	Mismatches
				Indels	0; Gaps
OY	1 TGCGCCGCTGTTCGGGCA	18			
DG	TGCCCGCTGTTCGGGCA	4138695			

RESULT 9
 US-09-103-840A-1/C
 : Sequence 1, Application US/09103840A
 : Patent No. 6294328
 : GENERAL INFORMATION:
 : APPLICANT: FLEISCHMAN, Robert D.
 : APPLICANT: WHITE, Owen R.
 : APPLICANT: FRASER, Claire M.
 : APPLICANT: VENTER, John C.
 : TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
 : TITLE OF INVENTION: TUBERCULOSIS
 : FILE REFERENCE: 24366-20007.00
 : CURRENT APPLICATION NUMBER: US/09/103.840A
 : CURRENT FILING DATE: 1998-06-24

```

: NUMBER OF SEQ ID NOS: 2
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 1
:
: LENGTH: 441529
:
: TYPE: DNA
:
: ORGANISM: Mycobacterium tuberculosis
:
: OTHER INFORMATION: H37Rv
US-09-103-840A-1

```

	Query Match	Score	DB	length
Best Local Similarity	91.1%;	16.4;	4;	4411529;
Matches	94.4%;	Pred. No. 18;		
Conservative	0;	Mismatches	1;	Indels 0;
				Gaps 0;
QY	1	TGGCGCCCTGTCTCGGCA	18	
DB	807416	TGGCGCCCTGTCTCGGCA	807399	

RESULT 10 09-452-239-35/C
 ; Sequence 35, Application US/09452239
 ; Patent No. 6465229
 ; GENERAL INFORMATION:
 ; APPLICANT: Rafalski, Antoni J.
 ; APPLICANT: Fader, Gary M.
 ; APPLICANT: Cahoon, Rebecca E.
 ; TITLE OF INVENTION: Plant Caateoy1-CoA O-Methyltransferase
 ; FILE REFERENCE: Bbl284 US NA
 ; CURRENT APPLICATION NUMBER: US/09/452,239
 ; CURRENT FILING DATE: 1999-12-01
 ; EARLIER APPLICATION NUMBER: 60/110,594
 ; EARLIER FILING DATE: 1998-December-02
 ; NUMBER OF SEQ ID NOS: 50
 ; SOFTWARE: Microsoft Office 97
 ; SEQ ID NO 35

```

1 TYPE: DNA
2 ORGANISM: Triticum aestivum
3 FEATURE:
4 NAME/KEY: unsure
5 LOCATION: (817)
6 FEATURE:
7 NAME/KEY: unsure
8 LOCATION: (826)
9 FEATURE:
10 NAME/KEY: unsure
11 LOCATION: (874)
12 FEATURE:
13 NAME/KEY: unsure
14 LOCATION: (891)
15 FEATURE:
16 NAME/KEY: unsure
17 LOCATION: (924)
18 FEATURE:
19 NAME/KEY: unsure
20 LOCATION: (934)
21 FEATURE:
22 NAME/KEY: unsure
23 LOCATION: (961)
24 FEATURE:
25 NAME/KEY: unsure
26 LOCATION: (970)..(971)
27 FEATURE:
28 NAME/KEY: unsure
29 LOCATION: (1012)
30 OS-09-452-239-35

```

Query Match	82.2%	Score 14.8	DB 4	Length 1018
Best Local Similarity	88.9%	Pred. No. 1.5e+02		
Matches 16	Conservative 0	Mismatches 2	Indels 0	Gaps 0
QY	1 TGGCCCCCTGTGCGGGA 18			

RESULT 14
US-09-199-637A-111
Sequence 111, Application US/09199637A
Patent No. 6355411
GENERAL INFORMATION:
APPLICANT: Ausubel, Frederick
APPLICANT: Goodman, Howard M.
APPLICANT: Rahne, Laurence G.
APPLICANT: Mahajan-Miklos, Shaila
APPLICANT: Cao, Hui
APPLICANT: Drenkard, Eliana
APPLICANT: Tsongalis, John
TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
TITLE OF INVENTION: SEQUENCES AND USES THEREOF
FILE REFERENCE: 00786/361002
CURRENT APPLICATION NUMBER: US/09/199,637A
CURRENT FILING DATE: 1998-11-25
PRIOR APPLICATION NUMBER: 60/066,517
PRIOR FILING DATE: 1997-11-25
NUMBER OF SEQ ID NOS: 437
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 111
LENGTH: 6909
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-199-637A-111

Query Match
Best Local Similarity 82.2%; Score 14.8; DB 4; Length 6909;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGGCGCGGTGTGCGGGA 18
|||||
DB 2627 TGGCGCGGTGTGCGGGA 2644

RESULT 15
US-08-611-757-14
Sequence 14, Application US/08611757
Patent No. 5859230
GENERAL INFORMATION:
APPLICANT: Kim, Jungsub P.
APPLICANT: Reyes, Gregory R.
APPLICANT: Wages, John
APPLICANT: Zhang-Keck, Zhen-Yang
APPLICANT: Young, Lavonne
TITLE OF INVENTION: No. 5859230-A/No. 5859230-B/No. 5859230-C/No. 5859230-D/No. 58
TITLE OF INVENTION: Agents and Molecular Cloning Thereof
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dehlinger & Associates
STREET: 350 Cambridge Avenue, Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/611,757
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/246,985
FILING DATE: 20-MAY-1994
APPLICATION NUMBER: US 025,396
FILING DATE: 24-FEB-1993

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/922,493
FILING DATE: 30-JUL-1992
ATTORNEY/AGENT INFORMATION:
NAME: Fabian, Gary R.
REGISTRATION NUMBER: 33,875
REFERENCE/DOCKET NUMBER: 4600-0201
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 324-0880
TELEFAX: (415) 324-0960
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 231 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: GB Clone 475-12-3
US-08-611-757-14

Query Match
Best Local Similarity 80.0%; Score 14.4; DB 2; Length 231;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 GCGCGGTGTGCGGGA 18
|||||
DB 131 GCGCGGTGTGCGGGA 146

Search completed: June 20, 2003, 23:25:44
Job time: 14.1792 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2003, 21:58:41 ; Search time 6.65094 Seconds
(without alignments)
3971.420 Million cell updates/sec

Title: US-09-965-553-2
Perfect score: 18
Sequence: 1 tggccgctgttcgggga 18

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1042519 seqs, 733713590 residues

Total number of hits satisfying chosen parameters: 2085038

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published_Applications_NA.*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCr_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCrUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18	100.0	18	10	US-09-965-553-2
2	18	100.0	27	10	US-09-965-553-39
3	18	100.0	9829	10	US-09-965-553-19
4	18	100.0	10482	10	US-09-965-553-23
5	18	100.0	12286	10	US-09-965-553-17
6	16.4	91.1	2000	9	US-09-938-842A-5131
7	16.4	91.1	2364	9	US-09-712-363-32
8	15.4	85.6	251	10	US-09-878-574-10644
9	15.4	85.6	262	10	US-09-878-574-12308
10	15.4	85.6	269	10	US-09-878-574-12035
11	15.4	85.6	269	10	US-09-878-574-13422
12	15.4	85.6	272	10	US-09-878-574-11753
13	15.4	85.6	273	10	US-09-878-574-15624
14	15.4	85.6	275	10	US-09-878-574-12670
15	15.4	85.6	275	10	US-09-878-574-14365
16	15.4	85.6	276	10	US-09-878-574-8192
17	15.4	85.6	285	10	US-09-878-574-12806
18	15.4	85.6	449	9	US-09-918-995-28499
19	15.4	85.6	1649	9	US-09-997-816-1

20	14.8	82.2	128	10	US-09-864-761-26007	Sequence 26007, A
21	14.8	82.2	254	10	US-09-923-876-1817	Sequence 1817, Ap
22	14.8	82.2	496	10	US-09-918-995-401	Sequence 401, App
23	14.8	82.2	561	10	US-09-864-761-9555	Sequence 9555, Ap
24	14.8	82.2	944	10	US-09-974-300-5723	Sequence 5723, Ap
25	14.8	82.2	1018	10	US-09-452-239-35	Sequence 35, App
26	14.8	82.2	1078	9	US-09-452-239-41	Sequence 41, App
27	14.8	82.2	1110	9	US-10-121-988-79	Sequence 79, App
28	14.8	82.2	1110	9	US-10-121-988-145	Sequence 145, App
29	14.8	82.2	1113	9	US-10-121-988-82	Sequence 82, App
30	14.8	82.2	1215	9	US-09-738-626-1258	Sequence 1258, Ap
31	14.8	82.2	3348	9	US-10-138-838-94	Sequence 94, App
32	14.8	82.2	3348	9	US-10-139-031-94	Sequence 94, App
33	14.8	82.2	3348	9	US-10-138-905-94	Sequence 94, App
34	14.8	82.2	3348	9	US-10-138-916-94	Sequence 94, App
35	14.8	82.2	3348	9	US-09-976-800-94	Sequence 94, App
36	14.8	82.2	6909	9	US-09-975-719-111	Sequence 111, App
37	14.8	82.2	3309400	9	US-09-738-626-1	Sequence 1, App
38	14.4	80.0	189	10	US-09-821-167-8	Sequence 8, App
39	14.4	80.0	249	10	US-09-878-574-15547	Sequence 15547, A
40	14.4	80.0	264	10	US-09-878-574-15236	Sequence 15236, A
41	14.4	80.0	269	10	US-09-878-574-10867	Sequence 10867, A
42	14.4	80.0	271	10	US-09-878-574-14193	Sequence 14193, A
43	14.4	80.0	303	10	US-09-878-574-15351	Sequence 15351, A
44	14.4	80.0	334	10	US-09-878-574-668	Sequence 668, App
45	14.4	80.0	342	10	US-09-878-574-2125	Sequence 2125, Ap

ALIGNMENTS

RESULT 1
US-09-965-553-2
; Sequence 2, Application US/09965553
; Patent No. US20020112259A1
; GENERAL INFORMATION:
; APPLICANT: Wright, David A.
; TITLE OF INVENTION: Plant Retroelements and Methods Related Thereto
; FILE REFERENCE: P-1065 ISURF Plant Retroelement
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 09/322,478
; PRIOR FILING DATE: 1999-05-28
; PRIOR APPLICATION NUMBER: 60/087125
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 18
; TYPE: DNA
; ORGANISM: Glycine max
US-09-965-553-2

Query Match 100.0%; Score 18; DB 10; Length 18;
Best Local Similarity 100.0%; Pred No. 5.2; Mismatches 0; Gaps 0;
Matches 18; Conservative 0; Indels 0;

QY 1 TGGCGCCGTTGTCGGGGA 18
|||||
DB 1 TGGCGCCGTTGTCGGGGA 18

RESULT 2
US-09-965-553-39
; Sequence 39, Application US/09965553
; Patent No. US20020112259A1
; GENERAL INFORMATION:
; APPLICANT: Wright, David A.
; APPLICANT: Voytas, Daniel F.
; TITLE OF INVENTION: Plant Retroelements and Methods Related Thereto
; FILE REFERENCE: P-1065 ISURF Plant Retroelement
; CURRENT APPLICATION NUMBER: US/09/965,553

```

; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 09/322,478
; PRIOR FILING DATE: 1999-05-28
; PRIOR APPLICATION NUMBER: 60/087125
; PRIOR FILING DATE: 1998-05-29
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 39
; LENGTH: 27
; TYPE: DNA
; ORGANISM: Glycine max
US-09-965-553-39
```

```

Query Match      100.0%; Score 18; DB 10; Length 27;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 TGGCGCCGTTGTGCGGGA 18
        |||||||
DB      5 TGGCGCCGTTGTGCGGGA 22
```

```

RESULT 3
US-09-965-553-19
; Sequence 19, Application US/09965553
; Patent No. US20020112259A1
; GENERAL INFORMATION:
; APPLICANT: Wright, David A.
; APPLICANT: Voytas, Daniel F.
; TITLE OF INVENTION: Plant Retroelements and Methods Related Thereto
; FILE REFERENCE: P-1065 ISURF Plant Retroelement
; CURRENT APPLICATION NUMBER: US/09/965,553
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 09/322,478
; PRIOR FILING DATE: 1999-05-28
; PRIOR APPLICATION NUMBER: 60/087125
; PRIOR FILING DATE: 1998-05-29
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 19
; LENGTH: 9829
; TYPE: DNA
; ORGANISM: Glycine max
US-09-965-553-19
```

```

Query Match      100.0%; Score 18; DB 10; Length 9829;
Best Local Similarity 100.0%; Pred. No. 3.3;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 TGGCGCCGTTGTGCGGGA 18
        |||||||
DB     1312 TGGCGCCGTTGTGCGGGA 1329
```

```

RESULT 4
US-09-965-553-23
; Sequence 23, Application US/09965553
; Patent No. US20020112259A1
; GENERAL INFORMATION:
; APPLICANT: Wright, David A.
; APPLICANT: Voytas, Daniel F.
; TITLE OF INVENTION: Plant Retroelements and Methods Related Thereto
; FILE REFERENCE: P-1065 ISURF Plant Retroelement
; CURRENT APPLICATION NUMBER: US/09/965,553
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 09/322,478
; PRIOR FILING DATE: 1999-05-28
; PRIOR APPLICATION NUMBER: 60/087125
; PRIOR FILING DATE: 1998-05-29
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 23
; LENGTH: 10482
```

```

; TYPE: DNA
; ORGANISM: Glycine max
US-09-965-553-23
```

```

Query Match      100.0%; Score 18; DB 10; Length 10482;
Best Local Similarity 100.0%; Pred. No. 3.2;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 TGGCGCCGTTGTGCGGGA 18
        |||||||
DB     1448 TGGCGCCGTTGTGCGGGA 1465
```

```

RESULT 5
US-09-965-553-17
; Sequence 17, Application US/09965553
; Patent No. US20020112259A1
; GENERAL INFORMATION:
; APPLICANT: Wright, David A.
; APPLICANT: Voytas, Daniel F.
; TITLE OF INVENTION: Plant Retroelements and Methods Related Thereto
; FILE REFERENCE: P-1065 ISURF Plant Retroelement
; CURRENT APPLICATION NUMBER: US/09/965,553
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 09/322,478
; PRIOR FILING DATE: 1999-05-28
; PRIOR APPLICATION NUMBER: 60/087125
; PRIOR FILING DATE: 1998-05-29
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 17
; LENGTH: 12286
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: plant
US-09-965-553-17
```

```

Query Match      100.0%; Score 18; DB 10; Length 12286;
Best Local Similarity 100.0%; Pred. No. 3.2;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 TGGCGCCGTTGTGCGGGA 18
        |||||||
DB     1312 TGGCGCCGTTGTGCGGGA 1329
```

```

RESULT 6
US-09-938-842A-5131
; Sequence 5131, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 5131
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
```


US-09-938-842A-5131

Query Match 91.1%; Score 16.4; DB 9; Length 2000;
Best Local Similarity 94.4%; Pred. No. 23;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 TGGCGCCGTTGTCGGGGA 18
|||||
DB 1883 TGGCGCCGTTGTCGGGGA 1900

RESULT 7

US-09-712-363-32/C
; Sequence 32, Application US/09712363
; Patent No. US20020164588A1
; GENERAL INFORMATION:
; APPLICANT: Eisenberg, David
; APPLICANT: Rotstein, Sergio H.
; APPLICANT: Marcotte, Edward M.
; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
; FILE REFERENCE: 07419-032001
; CURRENT APPLICATION NUMBER: US/09/712,363
; PRIOR FILING DATE: 2000-11-13
; PRIOR APPLICATION NUMBER: PCT/US00/02246
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/179,531
; PRIOR FILING DATE: 2000-02-01
; PRIOR APPLICATION NUMBER: 60/117,844
; PRIOR FILING DATE: 1999-01-29
; PRIOR APPLICATION NUMBER: 60/118,206
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: 60/126,593
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 60/134,093
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/134,092
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/165,124
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/165,086
; NUMBER OF SEQ ID NOS: 292
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 32
; LENGTH: 2364
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
US-09-712-363-32

Query Match 91.1%; Score 16.4; DB 9; Length 2364;
Best Local Similarity 94.4%; Pred. No. 23;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 TGGCGCCGTTGTCGGGGA 18
|||||
DB 1084 TGGCGCCGTTGTCGGGGA 1067

RESULT 8

US-09-878-574-10644/C
; Sequence 10644, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Thompson, Michael D.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(15401)B
; CURRENT APPLICATION NUMBER: US/09/878,574
; PRIOR FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/333,535

; PRIOR FILING DATE: 1999-06-14
; NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 10644
; LENGTH: 251
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: 700967905H1
US-09-878-574-10644

Query Match 85.6%; Score 15.4; DB 10; Length 251;
Best Local Similarity 94.1%; Pred. No. 87;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 2 GCGCGCCGTTGTCGGGGA 18
|||||
DB 243 GCGCGCCGTTGTCGGTGA 227

RESULT 9

US-09-878-574-12308/C
; Sequence 12308, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Thompson, Michael D.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(15401)B
; CURRENT APPLICATION NUMBER: US/09/878,574
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/333,535
; PRIOR FILING DATE: 1999-06-14
; NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 12308
; LENGTH: 262
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: 701065609H1
US-09-878-574-12308

Query Match 85.6%; Score 15.4; DB 10; Length 262;
Best Local Similarity 94.1%; Pred. No. 87;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 2 GCGCGCCGTTGTCGGGGA 18
|||||
DB 244 GCGCGCCGTTGTCGGTGA 228

RESULT 10

US-09-878-574-12035/C
; Sequence 12035, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Thompson, Michael D.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(15401)B
; CURRENT APPLICATION NUMBER: US/09/878,574
; PRIOR FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/333,535
; PRIOR FILING DATE: 1999-06-14
; NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 12035
; LENGTH: 269
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: 701065234H1
US-09-878-574-12035

Query Match 85.6%; Score 15.4; DB 10; Length 269;
Best Local Similarity 94.1%; Pred. No. 86;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 GGCGCCGTTGTCGGGGA 18
DB 243 GGCGCCGTTGTCGGTGA 227

RESULT 11
US-09-878-574-13422/C
Sequence 13422, Application US/09878574
Patent No. US20020110548A1

GENERAL INFORMATION:
APPLICANT: Byrum, Joseph R.
APPLICANT: La Rosa, Thomas J.
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
FILE REFERENCE: 38-21(15401)B
CURRENT APPLICATION NUMBER: US/09/878,574
CURRENT FILING DATE: 2001-12-21
PRIOR APPLICATION NUMBER: 09/333,535
PRIOR FILING DATE: 1999-06-14
NUMBER OF SEQ ID NOS: 15775
SEQ ID NO 13422
LENGTH: 269
TYPE: DNA
ORGANISM: Glycine max
OTHER INFORMATION: Clone ID: 701067081H1
US-09-878-574-13422

Query Match 85.6%; Score 15.4; DB 10; Length 269;
Best Local Similarity 94.1%; Pred. No. 86;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 GGCGCCGTTGTCGGGGA 18
DB 268 GGCGCCGTTGTCGGTGA 252

RESULT 12
US-09-878-574-11753/C
Sequence 11753, Application US/09878574
Patent No. US20020110548A1

GENERAL INFORMATION:
APPLICANT: Byrum, Joseph R.
APPLICANT: La Rosa, Thomas J.
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
FILE REFERENCE: 38-21(15401)B
CURRENT APPLICATION NUMBER: US/09/878,574
CURRENT FILING DATE: 2001-12-21
PRIOR APPLICATION NUMBER: 09/333,535
PRIOR FILING DATE: 1999-06-14
NUMBER OF SEQ ID NOS: 15775
SEQ ID NO 11753
LENGTH: 272
TYPE: DNA
ORGANISM: Glycine max
OTHER INFORMATION: Clone ID: 701064849H1
US-09-878-574-11753

Query Match 85.6%; Score 15.4; DB 10; Length 272;
Best Local Similarity 94.1%; Pred. No. 86;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 GGCGCCGTTGTCGGGGA 18
DB 226 GGCGCCGTTGTCGGTGA 210

RESULT 13
US-09-878-574-15624/C
Sequence 15624, Application US/09878574
Patent No. US20020110548A1

GENERAL INFORMATION:
APPLICANT: Byrum, Joseph R.
APPLICANT: La Rosa, Thomas J.
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
FILE REFERENCE: 38-21(15401)B
CURRENT APPLICATION NUMBER: US/09/878,574
CURRENT FILING DATE: 2001-12-21
PRIOR APPLICATION NUMBER: 09/333,535
PRIOR FILING DATE: 1999-06-14
NUMBER OF SEQ ID NOS: 15775
SEQ ID NO 15624
LENGTH: 273
TYPE: DNA
ORGANISM: Glycine max
OTHER INFORMATION: Clone ID: 701070324H1
US-09-878-574-15624

Query Match 85.6%; Score 15.4; DB 10; Length 273;
Best Local Similarity 94.1%; Pred. No. 86;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 GGCGCCGTTGTCGGGGA 18
DB 239 GGCGCCGTTGTCGGTGA 223

RESULT 14
US-09-878-574-12670/C
Sequence 12670, Application US/09878574
Patent No. US20020110548A1

GENERAL INFORMATION:
APPLICANT: Byrum, Joseph R.
APPLICANT: La Rosa, Thomas J.
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
FILE REFERENCE: 38-21(15401)B
CURRENT APPLICATION NUMBER: US/09/878,574
CURRENT FILING DATE: 2001-12-21
PRIOR APPLICATION NUMBER: 09/333,535
PRIOR FILING DATE: 1999-06-14
NUMBER OF SEQ ID NOS: 15775
SEQ ID NO 12670
LENGTH: 275
TYPE: DNA
ORGANISM: Glycine max
OTHER INFORMATION: Clone ID: 701066064H1
US-09-878-574-12670

Query Match 85.6%; Score 15.4; DB 10; Length 275;
Best Local Similarity 94.1%; Pred. No. 86;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 GGCGCCGTTGTCGGGGA 18
DB 235 GGCGCCGTTGTCGGTGA 219

RESULT 15
US-09-878-574-14365/C
Sequence 14365, Application US/09878574
Patent No. US20020110548A1

GENERAL INFORMATION:
APPLICANT: Byrum, Joseph R.
APPLICANT: La Rosa, Thomas J.
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with

```
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(15401)B
; CURRENT APPLICATION NUMBER: US/09/878,574
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/333,535
; PRIOR FILING DATE: 1999-06-14
; NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 14365
; LENGTH: 275
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: 701068668H1
; US-09-878-574-14365

Query Match      85.6%; Score 15.4; DB 10; Length 275;
Best Local Similarity 94.1%; Pred. No. 86;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      2 GCGCGCGTGTGCGGGA 18
        |||||||||||
Db      246 GCGCGCGTGTGCGTGA 230
```

Search completed: June 21, 2003, 00:34:40
Job time : 8.65094 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2003, 18:49:21 ; Search time 69.1132 Seconds
(without alignments)
4217.989 Million cell updates/sec

Title: US-09-965-553-2
Perfect score: 18
Sequence: 1 tggcgccgtgtcgggga 18

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_tod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18	100.0	250	17	AQ288973 nbxb0033H
2	18	100.0	552	17	BH731480 BOHZ016TF
3	18	100.0	624	17	BH545031 BOGCW07TF
4	18	100.0	668	17	BH600565 BOGDC03TF
5	18	100.0	678	17	BH214458 Gm_DMD001
6	18	100.0	690	17	BH484542 BOGWE45TF

7	18	100.0	694	17	BH725980 BOMMT13TF
8	18	100.0	697	13	BT642756 TZS973_T2
9	18	100.0	708	17	BH552125 BOHES25TF
10	18	100.0	720	17	BH445346 BOHNP83TF
11	18	100.0	722	17	BH424623 BOGEE39TF
12	18	100.0	741	17	BH141140 Gm_DMD001
13	18	100.0	747	17	BH505910 BOGDT64TF
14	18	100.0	768	17	BH590633 BOHXB52TF
15	18	100.0	770	17	BH423648 BOGMY06TF
16	18	100.0	771	17	BH700732 BOMNZ42TF
17	18	100.0	782	12	BG299408 HVMSEA002
18	18	100.0	789	17	BH724831 BOMLV90TF
19	18	100.0	796	17	BH514436 BOGUL71TF
20	18	100.0	805	17	BH531932 BOGRM94TF
21	18	100.0	809	17	BH420835 BOGCM07TF
22	18	100.0	819	17	BH454350 BOGBJ86TF
23	18	100.0	825	12	BF620501 HVMSEC002
24	18	100.0	828	17	BH583501 BOGHN35TF
25	18	100.0	848	17	BH451469 BOGXX39TF
26	18	100.0	858	17	BH698149 BOMNR68TF
27	18	100.0	864	17	BH246588 BOGAP59TF
28	17	94.4	719	17	BH591785 BOHAJ29TF
29	16.4	91.1	85	17	AZ922024 HRCOC2B09
30	16.4	91.1	147	10	AZ921883 HRCOC3E09
31	16.4	91.1	159	17	AM679402 WSL_24_CO
32	16.4	91.1	159	17	AM679402 WSL_24_CO
33	16.4	91.1	159	17	BH650606 BOGWE29TF
34	16.4	91.1	160	17	AZ922212 MKCOC2H07
35	16.4	91.1	160	17	BH501549 BOHMM57TF
36	16.4	91.1	231	17	AZ922191 MKCOC2D08
37	16.4	91.1	238	17	AL094958 Arabidops
38	16.4	91.1	282	17	BH436253 BOGHC70TF
39	16.4	91.1	288	17	BH744134 9332C02.D
40	16.4	91.1	291	17	BH705388 BOHVA74TF
41	16.4	91.1	300	17	BH745888 9273608.D
42	16.4	91.1	323	17	BH514295 BOGJO18TF
43	16.4	91.1	326	17	BH557816 BOGXT09TF
44	16.4	91.1	334	17	BH700667 BOMHO22TF
45	16.4	91.1	336	17	BH727785 BOHZN26TF

ALIGNMENTS

RESULT 1
LOCUS AQ288973 250 bp DNA linear GSS 03-DEC-1998
DEFINITION nbxb0033H14f CUGI Rice BAC Library Oryza sativa genomic clone
ACCESSION AQ288973.1 GI:3950419
VERSION AQ288973.1 GI:3950419
KEYWORDS GSS.
SOURCE Oryza sativa.
ORGANISM Oryza sativa.

REFERENCE 1 (bases 1 to 250)
AUTHORS Wing, R.A. and Dean, R.A.
TITLE A BAC End Sequencing Framework to Sequence the Rice Genome
JOURNAL Unpublished (1998)
COMMENT Contact: wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: TAATGAGCTCAGTATAGG
Class: BAC ends
High quality sequence stop: 117.
Location/Qualifiers L. 250

FEATURES
source

/organism="Oryza sativa"
 /strain="Japonica"
 /cultivar="Nipponbare"
 /db_xref="taxon:4530"
 /clone_1ib="CUGI Rice BAC library"
 /issue_type="Leaf"
 /lab_host="E. coli DH10B"
 /note="Vector: pBelobAC11; Site.1: HindIII; Site.2: HindIII; Rice is one of two most popular grains in the world. Half of the world population especially those inhabiting highly populated areas of the humid tropics and subtropics, rely on rice as their primary source of carbohydrate. Monocotyledonous rice is a diploid plant (2n=24) with a haploid genome equivalent of 431 Mbp (Arumuganathan and Earle, 1991). The relatively small genome of rice, three times larger than that of Arabidopsis, makes it suitable for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from Oryza sativa, Nipponbare variety. The library contains 36,864 clones with an average insert size of 128.5 Kb providing 10.9 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9 %. Two high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening."

BASE COUNT 88 a 49 c 59 g 54 t

ORIGIN

Query Match 100.0%; Score 18; DB 17; Length 250;
 Best Local Similarity 100.0%; Pred. No. 2e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TGGCGCCGTTGTCGGGA 18
 165 TGGCGCCGTTGTCGGGA 148

Db

RESULT 2
 BH731480
 LOCUS BOH2016F BO_2_3_KB Brassica oleracea genomic clone BOH2016, DNA
 DEFINITION
 accession BH731480 552 bp DNA linear GSS 20-FEB-2002
 version BH731480
 keywords GI:18836875
 source GSS.
 organism Brassica oleracea.
 reference Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eustosids II; Brassicales; Brassicaceae; Brassica.
 title 1 (bases 1 to 552)
 authors Town,C.D., Van Aken,S., Uterback,T. and Fraser,C.M.
 journal Whole genome shotgun sequencing of Brassica oleracea unpublished (2001)
 comment Other_GSSs: BOH2016TR
 contact: Chris Town
 TIGR
 9712 Medical Center Drive, Rockville, MD 20850, USA.
 Tel: 301-838-3523
 Fax: 301-838-0208
 Email: cdtown@tigr.org
 DNA is from a doubled haploid provided by Tom Osborn.
 Seq primer: TF
 Class: sheared ends.

FEATURES
 source
 1..552
 /organism="Brassica oleracea"
 /strain="T01000DH3"
 /db_xref="taxon:3712"
 /clone_1ib="BO_2_3_KB"

FEATURES
 source
 1..552
 /organism="Brassica oleracea"
 /strain="T01000DH3"
 /db_xref="taxon:3712"
 /clone_1ib="BOH2016"
 /clone_1ib="BO_2_3_KB"

/note="Vector: PHOS1; Site.1: BstXI; 2-3 kb sheared genomic DNA inserted into PHOS1 using BstXI linkers"
 BASE COUNT 154 a 111 c 97 g 190 t
 ORIGIN

Query Match 100.0%; Score 18; DB 17; Length 552;
 Best Local Similarity 100.0%; Pred. No. 2.5e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TGGCGCCGTTGTCGGGA 18
 364 TGGCGCCGTTGTCGGGA 381

Db

RESULT 3
 BH545031
 LOCUS BOGCW07TF BOGC Brassica oleracea genomic clone BOGCW07, DNA
 DEFINITION
 accession BH545031 624 bp DNA linear GSS 14-DEC-2001
 version BH545031
 keywords GI:17796812
 source GSS.
 organism Brassica oleracea.
 reference Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eustosids II; Brassicales; Brassicaceae; Brassica.
 title 1 (bases 1 to 624)
 authors Town,C.D., Van Aken,S., Uterback,T. and Fraser,C.M.
 journal Whole genome shotgun sequencing of Brassica oleracea unpublished (2001)
 comment Other_GSSs: BOGCW07TR
 contact: Chris Town
 TIGR
 9712 Medical Center Drive, Rockville, MD 20850, USA.
 Tel: 301-838-3523
 Fax: 301-838-0208
 Email: cdtown@tigr.org
 DNA is from a doubled haploid provided by Tom Osborn.
 Seq primer: TF
 Class: sheared ends.

FEATURES
 source
 1..624
 /organism="Brassica oleracea"
 /strain="T01000DH3"
 /db_xref="taxon:3712"
 /clone_1ib="BOGCW07"
 /clone_1ib="BOGC"
 /note="Vector: PHOS1; Site.1: BstXI; 2-3 kb sheared genomic DNA inserted into PHOS1 using BstXI linkers"

BASE COUNT 185 a 129 c 109 g 201 t

ORIGIN

Query Match 100.0%; Score 18; DB 17; Length 624;
 Best Local Similarity 100.0%; Pred. No. 2.6e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TGGCGCCGTTGTCGGGA 18
 542 TGGCGCCGTTGTCGGGA 559

Db

RESULT 4
 BH600565/c
 LOCUS BOGDG03TR BOGD Brassica oleracea genomic clone BOGDG03, DNA
 DEFINITION
 accession BH600565 668 bp DNA linear GSS 15-DEC-2001
 version BH600565
 keywords GI:17853011
 source GSS.
 organism Brassica oleracea.
 reference Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

REFERENCE
 TITLES
 JOURNAL
 COMMENT

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica. 1 (bases 1 to 668)
 Town, C.D., Van Aken, S., Utterback, T., and Fraser, C.M. Whole genome shotgun sequencing of Brassica oleracea unpublished (2001)
 Other GSSs: BOGD037F
 Contact: Chris Town

TIGR
 9712 Medical Center Drive, Rockville, MD 20850, USA.
 Tel: 301-838-3523
 Fax: 301-838-0208
 Email: cdtown@tigr.org
 DNA is from a doubled haploid provided by Tom Osborn.
 Seq primer: TR
 Class: sheared ends.

FEATURES
 source

1.668
 /organism="Brassica oleracea"
 /strain="T01000DH3"
 /db_xref="taxon:3712"
 /clone_lib="BOGD"
 /note="Vector: PHOS1, Site 1: BstXI; 2-3 kb sheared genomic DNA inserted into PHOS1 using BstXI linkers"

BASE COUNT
 186 a 146 c 144 g 192 t

Query Match
 Best Local Similarity 100.0%; Score 18; DB 17; Length 668;
 Pred. No. 2.7e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

ORIGIN
 1 TGGCGCCGTTGTCGGGGA 18
 ||||||||||||||||
 523 TGGCGCCGTTGTCGGGGA 506

RESULT 5
 BH214458
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

BH214458 678 bp DNA linear GSS 25-OCT-2001
 Gm_UMB001_098_A01.F UMN Soybean BAC Library (PECSBAC4 EcorI)
 Glycine max genomic clone Glycine max genomic clone
 Gm_UMB001_098_A01, DNA sequence.
 BH214458
 BH214458.1 GI:16428384
 GSS.
 soybean.
 Glycine max
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
 Glycine. 1 (bases 1 to 678)
 Marek, L.F., Paz, M., Darnelle, L., Hanson, N. and Shoemaker, R.C.
 BAC End sequences from a soybean genomic library (ISU)
 Unpublished (2000)
 Contact: Shoemaker Randy C
 Agronomy Department
 Iowa State University
 Ames, IA 50011-1010, USA
 Tel: 515 294 1205
 Fax: 515 294 2299
 Email: rcsshoelastate.edu
 This BAC identified by SSR Satt077. For more information, see
 Soybase at:
 http://genome.cornell.edu/cgi-bin/webace/webace?db=soybase. Please
 see as an authority for the mapping/naming: Cregan P.B., T. Jarvik,
 A.L. Bush, R.C. Shoemaker, K.G. Iark, A.L. Kahler, N. Kaya, T.T.
 Vantcal, D.G. Lohnes, J. Chung, and J.E. Specht. 1999a. An
 integrated genetic linkage map of the soybean genome. Crop Sci.
 39:1464-1490.
 Seq primer: M13F
 Class: BAC ends.

FEATURES
 source

Location/Qualifiers
 1.678
 /organism="Glycine max"
 /cultivar="Faribault"
 /db_xref="taxon:3847"
 /clone="Gm_UMB001_098_A01"
 /clone_lib="UMN Soybean BAC Library (PECSBAC4 EcorI)
 Glycine max genomic clone"
 /tissue_type="cotyledon leaves"
 /dev_stage="cotyledon"
 /note="Vector: PECSBAC4; The UMN BAC library (Danesh et al
 Theor. Appl. Genet. 96:196, 1998) was constructed using
 the Eco RI site of PECSBAC4. The library consists of 72
 to 7 haploid clones with an average insert size of 120 Kb, equal
 is done by hybridization of high-density colony filters
 and/or PCR amplification of DNA pools. Four high density
 filters, each containing 18,432 clones (doubly spotted),
 represent the whole library for colony screening"

BASE COUNT
 180 a 92 c 80 g 301 t

Query Match
 Best Local Similarity 100.0%; Score 18; DB 17; Length 678;
 Pred. No. 2.7e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

ORIGIN
 1 TGGCGCCGTTGTCGGGGA 18
 ||||||||||||||||
 539 TGGCGCCGTTGTCGGGGA 556

RESULT 6
 BH484542/c
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

BH484542 690 bp DNA linear GSS 13-DEC-2001
 BOGWE45TF BOGW Brassica oleracea genomic clone BOGWE45, DNA
 sequence.
 BH484542
 BH484542.1 GI:17692646
 GSS.
 Brassica oleracea.
 Brassica oleracea
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica. 1
 (bases 1 to 690)
 Town, C.D., Van Aken, S., Utterback, T., and Fraser, C.M.
 Whole genome shotgun sequencing of Brassica oleracea
 Unpublished (2001)
 Other GSSs: BOGWE45TF
 Contact: Chris Town

REFERENCE
 TITLES
 JOURNAL
 COMMENT

9712 Medical Center Drive, Rockville, MD 20850, USA.
 Tel: 301-838-3523
 Fax: 301-838-0208
 Email: cdtown@tigr.org
 DNA is from a doubled haploid provided by Tom Osborn.
 Seq primer: TF
 Class: sheared ends.

FEATURES
 source

1.690
 /organism="Brassica oleracea"
 /strain="T01000DH3"
 /db_xref="taxon:3712"
 /clone="BOGWE45"
 /clone_lib="BOGW"
 /note="Vector: PHOS1, Site 1: BstXI; 2-3 kb sheared genomic DNA inserted into PHOS1 using BstXI linkers"

BASE COUNT
 221 a 139 c 140 g 180 t

Query Match
 Best Local Similarity 100.0%; Score 18; DB 17; Length 690;
 Pred. No. 2.7e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGCGCCGTTGCGGGA 18
 |||
 Db 280 TGGCGCCGTTGCGGGA 263

RESULT 7
 BH725980 694 bp DNA linear GSS 20-FEB-2002
 LOCUS BOMMT13R_BO_2_3_KB Brassica oleracea genomic clone BOMMT13, DNA
 DEFINITION
 ACCESSION BH725980
 VERSION BH725980.1 GI:18831375
 KEYWORDS GSS.
 SOURCE Brassica oleracea.
 ORGANISM Brassica oleracea.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eustosids II; Brassicales; Brassicaceae; Brassica.
 1 (bases 1 to 694)
 Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
 Whole genome shotgun sequencing of Brassica oleracea
 Unpublished (2001)
 Other_GSSs: BOMMT13TF
 COMMENT
 TIGR
 9712 Medical Center Drive, Rockville, MD 20850, USA.
 Tel: 301-838-3523
 Fax: 301-838-0208
 Email: cdtown@tigr.org
 DNA is from a doubled haploid provided by Tom Osborn.
 Seq primer: TR
 Class: sheared ends.

FEATURES
 source
 1..694
 Location/Qualifiers
 /organism="Brassica oleracea"
 /strain="TO1000DH3"
 /db_xref="taxon:3712"
 /clone="BOMMT13"
 /clone_1lb="BO_2_3_KB"
 /note="Vector: PHOS1; Site_1: BstXI; 2-3 kb sheared
 genomic DNA inserted into PHOS1 using BstXI linkers"

BASE COUNT 188 a 182 c 139 g 185 t

ORIGIN
 Query Match 100.0%; Score 18; DB 17; Length 694;
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGCGCCGTTGCGGGA 18
 |||
 Db 133 TGGCGCCGTTGCGGGA 150

RESULT 8
 BI642756 697 bp mRNA linear EST 10-SEP-2001
 LOCUS T2S973 T2S (Sapwood-heartwood transition zone of black locust -
 DEFINITION Summer) Robinia pseudoacacia cDNA, mRNA sequence.
 ACCESSION BI642756
 VERSION BI642756.1 GI:15544966
 KEYWORDS EST.
 SOURCE Robinia pseudoacacia.
 ORGANISM Robinia pseudoacacia.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eustosids I; Fabales; Fabaceae; Papilionoideae; Robinieae;
 Robinia.
 1 (bases 1 to 697)
 Han,K.-H., Yang,J., Park,S., Paule,C.R., Kapur,V., Retzel,E.F.,
 Kamdem,D.P. and Keathley,D.E.
 Analysis of gene expression patterns in trunk wood of a mature
 black locust (Robinia pseudoacacia)

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

JOURNAL
 COMMENT Unpublished (2002)
 Contact: Kyung-Hwan Han
 Department of Forestry
 Michigan State University
 126 Natural Resources, East Lansing, MI 48824-1222, USA
 Tel: 517 353 4751
 Fax: 517 432 1143
 Email: hanky@msu.edu.

FEATURES
 source
 1..697
 Location/Qualifiers
 /organism="Robinia pseudoacacia"
 /db_xref="taxon:35938"
 /clone_1lb="T2S (Sapwood-heartwood transition zone of
 black locust - Summer)"
 /tissue_type="sapwood-heartwood transition zone"
 /dev_stage="mature tree"
 /note="Vector: lambda Triplex; Site_1: Sfi IA; Site_2: Sfi
 IB; The cDNA library was made from the sapwood-heartwood
 transition zone of a mature black locust tree collected in
 Michigan in late July." 123 c 94 g 208 t

BASE COUNT 272 a 123 c 94 g 208 t

ORIGIN
 Query Match 100.0%; Score 18; DB 13; Length 697;
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGCGCCGTTGCGGGA 18
 |||
 Db 25 TGGCGCCGTTGCGGGA 8

RESULT 9
 BH552125 708 bp DNA linear GSS 14-DEC-2001
 LOCUS BOHE525TF BOHE Brassica oleracea genomic clone BOHE525, DNA
 DEFINITION
 ACCESSION BH552125
 VERSION BH552125.1 GI:17803905
 KEYWORDS GSS.
 SOURCE Brassica oleracea.
 ORGANISM Brassica oleracea.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eustosids II; Brassicales; Brassicaceae; Brassica.
 1 (bases 1 to 708)
 Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
 Whole genome shotgun sequencing of Brassica oleracea
 Unpublished (2001)
 Other_GSSs: BOHE525TR
 COMMENT
 TIGR
 9712 Medical Center Drive, Rockville, MD 20850, USA.
 Tel: 301-838-3523
 Fax: 301-838-0208
 Email: cdtown@tigr.org
 DNA is from a doubled haploid provided by Tom Osborn.
 Seq primer: TF
 Class: sheared ends.

FEATURES
 source
 1..708
 Location/Qualifiers
 /organism="Brassica oleracea"
 /strain="TO1000DH3"
 /db_xref="taxon:3712"
 /clone="BOHE525"
 /clone_1lb="BOHE"
 /note="Vector: PHOS1; Site_1: BstXI; 2-3 kb sheared
 genomic DNA inserted into PHOS1 using BstXI linkers"

BASE COUNT 191 a 181 c 138 g 198 t

ORIGIN
 Query Match 100.0%; Score 18; DB 17; Length 708;
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;


```

/organism="Glycine max"
/cultivar="Fairbault"
/db_xref="taxon:3847"
/clone.lib="UNM Soybean BAC Library (PECSBAC4 ECORI)
glycine max genomic clone"
/tissue_type="cotyledon leaves"

```

```

/dev-stage="cotyledon"
/Note="Vector: pECSBAC4; The UMN BAC library (Danesh et al
Theor. Appl. Genet. 96:196, 1998) was constructed using
the Eco RI site of pECSBAC4. The library consists of 72
,960 clones with an average insert size of 120 kb, equal
to 7 haploid genome equivalents. Screening of the library
is done by hybridization of high-density colony filters
and/or PCR amplification of DNA pools. Four high density
filters, each containing 18,432 clones (doubly spotted),
represent the whole library for colony screening"

BASE COUNT      200 a      97 c      82 g      333 t
ORIGIN

Query Match      100.0%; Score 18; DB 17; Length 741;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TGGCGCCGTTGTGCGGGA 18
        |||||||
Db      548 TGGCGCCGTTGTGCGGGA 565

RESULT 13
BH505910      747 bp      DNA      linear      GSS 13-DEC-2001
DEFINITION    BOCGT64TR BOCB Brassica oleracea genomic clone BOCGT64, DNA
sequence.
ACCESSION     BH505910
VERSION       BH505910.1 GI:17714007
KEYWORDS      GSS.
SOURCE        Brassica oleracea.
ORGANISM      Brassica oleracea.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE     1 (bases 1 to 747)
AUTHORS       Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
TITLE         Whole genome shotgun sequencing of Brassica oleracea
JOURNAL       Unpublished (2001)
COMMENT       Other_GSSs: BOCGT64TF
               Contact: Chris Town
               TIGR
               9712 Medical Center Drive, Rockville, MD 20850, USA.
               Tel: 301-838-3523
               Fax: 301-838-0208
               Email: cdtown@tigr.org
               DNA is from a doubled haploid provided by Tom Osborn.
               Seq primer: TR
               Class: sheared ends.
               Location/Qualifiers
                 1..747
                 /organism="Brassica oleracea"
                 /strain="TO1000DH3"
                 /db_xref="taxon:3712"
                 /clone="BOGDT64"
                 /clone_id="BOGD"
                 /note="Vector: PHOS1; Site 1: BstXI; 2-3 kb sheared
                 genomic DNA inserted into PHOS1 using BstXI linkers"

BASE COUNT      214 a      188 c      136 g      209 t
ORIGIN

Query Match      100.0%; Score 18; DB 17; Length 747;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TGGCGCCGTTGTGCGGGA 18
        |||||||
Db      233 TGGCGCCGTTGTGCGGGA 250

RESULT 14
BH590633/c      768 bp      DNA      linear      GSS 15-DEC-2001
LOCUS

```

```

DEFINITION    BOHB252TR BOHB Brassica oleracea genomic clone BOHB252, DNA
sequence.
ACCESSION     BH590633
VERSION       BH590633.1 GI:17843085
KEYWORDS      GSS.
SOURCE        Brassica oleracea.
ORGANISM      Brassica oleracea.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE     1 (bases 1 to 768)
AUTHORS       Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
TITLE         Whole genome shotgun sequencing of Brassica oleracea
JOURNAL       Unpublished (2001)
COMMENT       Other_GSSs: BOHB252TR
               Contact: Chris Town
               TIGR
               9712 Medical Center Drive, Rockville, MD 20850, USA.
               Tel: 301-838-3523
               Fax: 301-838-0208
               Email: cdtown@tigr.org
               DNA is from a doubled haploid provided by Tom Osborn.
               Seq primer: TR
               Class: sheared ends.
               Location/Qualifiers
                 1..768
                 /organism="Brassica oleracea"
                 /strain="TO1000DH3"
                 /db_xref="taxon:3712"
                 /clone="BOHB252"
                 /clone_id="BOHB"
                 /note="Vector: PHOS1; Site 1: BstXI; 2-3 kb sheared
                 genomic DNA inserted into PHOS1 using BstXI linkers"

BASE COUNT      207 a      169 c      179 g      213 t
ORIGIN

Query Match      100.0%; Score 18; DB 17; Length 768;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TGGCGCCGTTGTGCGGGA 18
        |||||||
Db      651 TGGCGCCGTTGTGCGGGA 634

RESULT 15
BH423648/c      770 bp      DNA      linear      GSS 12-DEC-2001
LOCUS          BOCGW06TR BOCW Brassica oleracea genomic clone BOCGW06, DNA
sequence.
ACCESSION     BH423648
VERSION       BH423648.1 GI:17609376
KEYWORDS      GSS.
SOURCE        Brassica oleracea.
ORGANISM      Brassica oleracea.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE     1 (bases 1 to 770)
AUTHORS       Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
TITLE         Whole genome shotgun sequencing of Brassica oleracea
JOURNAL       Unpublished (2001)
COMMENT       Other_GSSs: BOCGW06TF
               Contact: Chris Town
               TIGR
               9712 Medical Center Drive, Rockville, MD 20850, USA.
               Tel: 301-838-3523
               Fax: 301-838-0208
               Email: cdtown@tigr.org
               DNA is from a doubled haploid provided by Tom Osborn.
               Seq primer: TR
               Class: sheared ends.
               Location/Qualifiers

```

```
source          1. 770
                 /organism="Brassica oleracea"
                 /strain="T0100DH3"
                 /db_xref="taxon:3712"
                 /clone="BOGWY06"
                 /clone_1lb="BOGW"
                 /note="Vector: PHOS1; Site_1: BstXI; 2-3 kb sheared
                 genomic DNA inserted into PHOS1 using BstXI linkers"
BASE COUNT      220 a 130 c 167 g 253 t
ORIGIN

Query Match     100.0%; Score 18; DB 17; Length 770;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 TGGCGCGGTGTGCGGGA 18
    ||||||||||||||||
Db 526 TGGCGCGGTGTGCGGGA 509
```

Search completed: June 20, 2003, 22:39:16
Job time : 73.1132 secs

THIS PAGE BLANK (USPTO)